Knowledge, methods and the impossibility of error

DUTANT, Julien

Abstract
This work defends method infallibilism, the thesis that propositional knowledge is belief based on an infallible method. The view belongs to the family of modal analyses of knowledge that have been developed in the last fifty years. We put these analyses in a historical perspective and characterise their form and variants rigorously. Method infallibilism is defended as the simplest modal analysis that satisfies two constraints needed to avoid the Gettier problem and what we call the necessary truth problem. We implement the view in a new formal model for knowledge and a contextualist account of knowledge attributions, though we leave open whether the latter should be endorsed.

Reference


URN : urn:nbn:ch:unige-235742
DOI : 10.13097/archive-ouverte/unige:23574
Knowledge, Methods and the Impossibility of Error

Julien Dutant

Thèse de doctorat ès lettres

Directeur de thèse
Professeur Pascal Engel, Université de Genève

Président du Jury
Professeur Kevin Mulligan, Université de Genève

2010
version révisée
Cette thèse a été défendue le 22 juin 2010 devant un jury constitué d’Igor Douven (KU Louvain), Duncan Pritchard (Edimbourg), François Récanati (Institut Jean Nicod, CNRS/ENS/EHESS), Ernest Sosa (Rutgers), Timothy Williamson (Oxford) et Kevin Mulligan (Genève, président). Je remercie chaleureusement les membres du jury pour leurs commentaires détaillés.

La thèse a été révisée après la soutenance pour corriger diverses coquilles et répondre à des difficultés soulevées par le jury. Les révisions sont indiquées comme telles, à l’exception des coquilles et autres changements mineurs. Seul l’appendice A a été substantiellement réécrit afin de se dispenser de la Présupposition de Limite de Lewis et afin de présenter plus fidèlement la sémantique de Nozick.

This thesis has been defended on June 22nd 2010 before a committee consisting of Igor Douven (KU Leuven), Duncan Pritchard (Edinburgh), François Récanati (Institut Jean Nicod, CNRS/ENS/EHESS), Ernest Sosa (Rutgers), Timothy Williamson (Oxford) and Kevin Mulligan (Geneva, president). I warmly thank the members of the committee for their detailed comments.

The thesis has been revised after the viva to remove a number of typos and to address some issues pointed out by the jury. Minor changes aside, revisions are indicated as such. Only Appendix A has been substantially rewritten to avoid relying on the Limit Assumption and to spell out more accurately Nozick’s semantics.
Short Contents

1 Introduction
   1.1 Method infallibilism and conceptual analysis ........ 1
   1.2 What methods are ................................................. 4
   1.3 Methods infallibilism and the theory of knowledge ...... 6
   1.4 Overview ................................................................. 8

I History of Epistemology .................................................. 13

2 The Legend of the Justified True Belief Analysis ............... 15
   2.1 The Legend and its puzzles ............................... 19
   2.2 Ur-foundationalism .................................................... 31
   2.3 Historical illustrations ............................................. 46
   2.4 Hume’s predicament, probabilism and idealism .......... 67
   2.5 The Gettier problem and the internalism/externalism divide 84
   2.6 Conclusion ................................................................. 95
### II Epistemology

#### 3 Modal Requirements
- 3.1 Conditions and requirements .......................... 103
- 3.2 Motivation for modal requirements .................. 109
- 3.3 The avoidance of error ................................. 116
- 3.4 Simple modal requirements ............................ 120
- 3.5 Parametric modal requirements ....................... 133
- 3.6 Reliability requirements ............................... 139
- 3.7 Double modal requirements ........................... 143
- 3.8 Discussion of the results .............................. 146

#### 4 Methods infallibilism
- 4.1 The Method constraint .................................. 155
- 4.2 The Infallibility constraint .............................. 171
- 4.3 Objections to necessity .................................. 183
- 4.4 Objections to sufficiency ............................... 207
- 4.5 Conclusion .............................................. 221

### III Formal Epistemology

#### 5 Methods-based Models
- 5.1 Introduction ............................................ 226
- 5.2 Methods ................................................. 231
- 5.3 Applications ............................................ 238
- 5.4 Methods models ....................................... 245
- 5.5 Results .................................................. 253
- 5.6 Conclusion .............................................. 270

### IV Semantics

#### 6 Modal Epistemic Contextualism
- 6.1 Epistemic contextualism ............................... 277
C.2 Comparison with neighbourhood models . . . . . . . . . . . . . 415
C.3 Counterexamples to M, N, K and 4 . . . . . . . . . . . . . . . . . . 420
C.4 Further exploration of Reasoning methods . . . . . . . . . . . . 423
C.5 Belief, knowledge and information . . . . . . . . . . . . . . . . . . 426

D Safety, Closure and Knowledge of the Past 431
D.1 Time-relative safety . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 431
D.2 Safety and knowledge of the past . . . . . . . . . . . . . . . . . . 433
D.3 Time-relative safety and deductive closure . . . . . . . . . . . . 438
D.4 Generalisation and ways out . . . . . . . . . . . . . . . . . . . . . . . 442

E Uncentred possibilities of error 447
E.1 Problems for subject-centring . . . . . . . . . . . . . . . . . . . . . . . 447
E.2 Uncentred method safety . . . . . . . . . . . . . . . . . . . . . . . . . 451
E.3 An objection and reply . . . . . . . . . . . . . . . . . . . . . . . . . . . 453

List of Symbols 459
List of Tables 463
List of Figures 465
Bibliography 467
Detailed Contents 493
This has been a long, eventful and enjoyable journey, and I almost regret that it is over. It started almost a decade ago, when I undertook a second year Master with Pascal Engel and spent one year as a visiting student in Oxford. I knew next to nothing of contemporary analytic philosophy, academics or the academic life then. I have not been disappointed. I would like to express my gratitude to the many people who helped and accompanied me along the way.

First and foremost, thanks to Pascal Engel for quietly leading me to much more interesting areas of philosophy than the ones I was originally attracted to. His supervision, encouragement, constant support, tolerance and trust have been invaluable. I also owe special thanks to François Récanati for his supervision, support, and a decisive orientation towards the topic of epistemic contextualism. Three other philosophers have deeply influenced the present work through their writings and discussion: Ernest Sosa, whose *Knowledge in Perspective* converted me early on to externalism, Timothy Williamson, whose *Knowledge and its Limits* inspired many themes in this work, and John Hawthorne, whose *Knowledge and Lotteries* has shaped my thinking on safety and epistemic contextualism. I am grateful to them. I also thank John Hawthorne for reading and commenting drafts of parts I and III and appendix D.

I have been lucky enough to prepare this work in two of the best philosophical homes there are in Europe: the Institut Jean Nicod in Paris and the Department of Philosophy in Geneva. I would like to thank the people I met at or around each institution for their philosophical stimulation and friendship: in Paris, Maud Barriquand, Nicolas Baumard,
Sarah Bendaoud, Alexandre Billon, Dennis Bonnay, Philippe de Brabant, Roberto Casati, Geneviève Choquette, Florian Cova, Mikaël Cozic, Camelia Dascalu, Jérôme Dokic, Christine van Geen, Emmanuelle Glon, Bastien Guerry, Marie Guillot, Benoît Hardy-Vallée, Christophe Heintz, Pierre Jacob, Max Kistler, Stéphane Lemaire, Pascal Ludwig, Edouard Machery, Olivier Morin, Karim N’Diaye, David Nicolas, Cédric Paternote, Elisabeth Pacherie, Jérôme Pelletier, Nausicaa Pouscoulous, Joëlle Proust, Philippe Schlenker, Benjamin Spector, Dan Sperber, Isidora Stojanovic, Benjamin Sylvand, Dario Taraborelli, Claudine Tiercelin, Anne Tüscher, Philippe Vellozzo, Frédérique de Vignemont and Neftali Villanueva-Fernandez; in Geneva, Lausanne and Fribourg, Bernard Baertschi, Otto Bruun, Christine Clavien, Ariel Cecchi, Laurent Cesalli, Fabrice Correia, Julien Deonna, Fabian Dorsch, Jacques Duparc, Michael Esfeld, Alessandro Facchini, Davide Fassio, Akiko Frischhut, Amanda Garcia, Ghislain Guigon, Samia Hurst, Dominique Janain, Philipp Keller, Federico Lauria, Jessica Leech, Arturs Logins, Jean-Roch Lauper, Olivier Massin, Anne Meylan, Kevin Mulligan, Martine Nida-Rümelin, Jessica Pallie, Alain Pé-Curto, Graham Peebles, Andreas Schmidhauser, Luc Schneider, Gianfranco Soldati, Johannes Stern, Selja Seppälä, Christian Sachse, Gian-Andri Toendury, Fabrice Teroni and Christian Woldemar-Bach. Special thanks to Paul Egré for much encouragement, inspiration and feedback and to Damian Justo for endless philosophical conversations. Marie-Claude Lorne should have been on this list; the memory of her intensity and courage has not left me.

I also owe much of my philosophical education to a year as a visiting student at Oxford in 2000–1 (especially Timothy Williamson’s and Dorothy Edgington’s philosophy of language seminar), a year as a M.A. student at University College London in 2001–2 (especially Mike Martin’s, Barry C. Smith’s and Scott Sturgeon’s seminars and Tim Crane’s and José Zalabardo’s supervision), a second year as a visiting student in Oxford in 2009–10 and to several summer schools (NYU - Paris IV - Bologna - Parma school on The a priori organised by Paolo Leonardi, Florence, 2003, Workshop on Jackendoff and school on Reference to Objects.

My students at the University of Paris-IV and the University of Geneva have made it worthwhile to be a philosopher. I would like to thank them for their curiosity and enthusiasm. Special thanks are due to the first to experiment the discussion seminar in Geneva: Pablo Carnino, Gabriel Carnino, Rodrigo Cortesão, Lorcan Meyfarth, Alan Monnat, and Stefan Heeb, and to David Fürrer and François Jaquet.

I have learnt a lot on the topics of this work from short and long discussions with Johan van Benthem, Michael Blome-Tillman, Berit Brogaard, J. Adam Carter, Dylan Dodd, Igor Douven, Manolo Garcia-Carpintero, Mikkel Gerken, Paul Gochet, Vincent Hendricks, Klemens Kappel, Chris Kelp, Angelika Kratzer, Mike Martin, Maria Lasonen-Aarnio, Frank Li-horeau, Conor McHugh, Erik Olsson, Martine Pécharman, Duncan Pritchard, Olivier Roy, Jonathan Shaheen, Barry C. Smith, Robert Stalnaker, Jason Stanley, Scott Sturgeon, Paula Sweeney, Jonas de Vuyst, Lee Walters, Andreas Witzel, José Zalabardo and Elia Zardini.

Ihady am also grateful to many colleagues and friends who have played a small or important in my life over these years: Daniel Andler, Hady Ba, Brian Ball, Corine Besson, Adrian Briciu, Justin Broackes, Reinaldo Bernal Velasquez, Martijn Blaauw, Delphine Blitman, Gregory Bochner, Sacha Bourgeois-Gironde, Cédric Brun, Josée Brunet, Stéphanie Burette, Oscar Cabaco, Benjamin Caraco, Nicolas Charles, Fabien Chareix, Jean-Marie Chevalier, Yannick Chin, Roberto Ciuni, Eros Corazza, Ar-mand David, Ophelia Deroy, Trent Dougherty, Jacques Dubucs, Santiago Etcheverri, Cédric Eyssette, Ekain Garmendia, Benoît Gaultier, Ludivine Godbille, Julie Higonnet, Mikhail Kissine, Max Kölbl, Julia Langkau,
Acknowledgements


Un grand merci aussi à ma famille, mon père Joël pour son ambition, mon frère Arnaud pour son intelligence, mes tantes Marie-Thérèse et Martine pour leur gentillesse et leur soutien, et tout particulièrement à ma mère Annie pour sa confiance inébranlable.

I am deeply grateful to Sanna Hirvonen for much support, great times, countless philosophical discussions and being so awesome.
The main ideas of Part I of the thesis have been published in my “Pourquoi le problème de Gettier est-il si important?”, in P. Ducray (ed.), *Revue Klesis*, 9 :2008, 64–104 and *Qu’est-ce que la connaissance?* Paris, Vrin, 2010, and have been presented at the Maison Française d’Oxford, Feb 8th 2010. Thanks to the audience and to Pascal Engel, Kevin Mulligan, John Hawthorne and Timothy Williamson for useful discussions of the topic.


Successive versions of Part III have been presented at Geneva (*PALMYR VIII, Paris-Amsterdam-Switzerland logic meeting of young researchers, may 8-9*th 2009), Palmyr workshop), Lausanne (*Decision, Games and Logic workshop 09, June 15-17*th 2009), Nancy (*Epistemology, Context, Formalism workshop, Archives Poincaré, Nov 12-14*th 2009). I have much benefited from comments form the audiences in each occasion. I am particularly indebted to two anonymous referees for detailed comments. Thanks are also due to Paul Egré, Elia Zardini, Jonathan Shaheen, Davide Fassio, Fabrice Correia, Olivier Roy, Andreas Witzel and Vincent Hendricks for much-needed help and encouragement.

Appendix D stems out of a paper presented at Edinburgh (Virtue Epistemology workshop, Apr 15th) and Geneva (Congress of the Société Francophone de Philosophie Analytique (SoPhA), Sept 2–6th, 2009 and PhilEAs talks, Nov 2th, 2009). Thanks to the audiences and to John Hawthorne and to Maria Lasonen-Aarnio for discussion.

Appendix E has not been presented nor published. Thanks to Sanna Hirvonen for a detailed discussion of it.

This work has been supported by an Allocataire-Moniteur position at the University of Paris IV (2003–6), by an Assistant position at the Department of Philosophy in Geneva (2006–9), and by the Swiss National Science Foundation project Connaissance, Raison et Normes (2007–10) and by a Swiss National Foundation Young Researcher grant (grant PBGEP1-125972, 2009-10). I am grateful to these institutions for their support. Thanks also to the Maison Française d’Oxford for providing me a ideal work environment in 2009–10.
Abstract

This work defends *method infallibilism*, the thesis that propositional knowledge is belief based on an infallible method. The view belongs to the family of modal analyses of knowledge that have been developed in the last fifty years. We put these analyses in a historical perspective and characterise their form and variants rigorously. Method infallibilism is defended as the simplest modal analysis that satisfies two constraints needed to avoid the Gettier problem and what we call the necessary truth problem. We implement the view in a new formal model for knowledge and a contextualist account of knowledge attributions, though we leave open whether the latter should be endorsed.

*Part I* puts method infallibilism in a historical perspective. We argue that knowledge was traditionally thought of as belief bearing a *discernible and infallible mark of truth*. The view, which we label *Ur-foundationalism*, is illustrated by Stoic epistemology. It ultimately leads to Hume’s scepticism, since few, if any, of our beliefs bear such marks. Contemporary internalists (typically) maintain the discernibility component while giving up infallibilism: this is the justified true belief analysis that Gettier refuted. Externalists instead give up discernibility while (typically) maintaining infallibility. Our historical picture explains the recent appearance of the Gettier problem, and goes against the widespread idea that the justified true belief analysis was the traditional one. It provides preliminary motivation for the externalist infallibilism we defend.

*Part II* presents the case for method infallibilism. We call *modal requirement* a statement of a condition for knowledge that a case satisfies if false belief was in some relevant sense impossible in that case. We show that
most recent modal accounts of knowledge are variants on that general form and set out their key structural features. We argue that if a modal account of knowledge is to be provided, it should be methods-based, to avoid trivialisation in the case of beliefs in necessary truths, and infallibilist, to avoid Gettier problems. We complete our case for method infallibilism by defusing objections to the necessity and sufficiency of method infallibility of knowledge. In particular, the generality problem is avoided by identifying methods in terms of knowledge as epistemically relevant classes of belief cases. Our defence is conditional on the desirability of a modal account of knowledge, in favour of which we provide some methodological considerations.

Part III implements method infallibilism formally. Methods are represented as functions from worlds and sets of premises to sets of conclusions; beliefs are defined as conclusions reached from the empty set of premises. We introduce an accessibility relation for alethic possibility, and define knowledge as belief based on a method that only yields true beliefs at accessible worlds. With propositions coarsely defined as sets of possible worlds, the resulting semantics is a significant extension of Scott’s and Montague’s neighbourhood semantics. In general it only validates the claim that knowledge entails truth and belief, but we show that the full S5 axioms of epistemic logic can be derived from a series of natural idealisations of agents. The models provide an enlightening perspective on why and when the S5 axioms hold and a more natural representation of epistemic notions like inductive knowledge and Gettier cases than standard Hintikka models.

Part IV links methods infallibilism with ordinary knowledge attributions and their semantics. We argue that a notion of real possibility is reflected in our use of circumstantial modals and subjunctive conditionals. Given some intuitive connections between the latter and knowledge ascriptions, we defend a semantics for “know” according to which knowledge requires avoidance of real possibilities of error. Standard treatments of circumstantial modals and subjective conditionals suggest that real possibility is a context-sensitive matter. The semantics thus yields a
version of contextualism about knowledge ascriptions. We show that it avoids a number of problems that DeRose’s and Lewis’s versions face — some of which not having been pointed out in the literature yet, such as the fact that Lewis’s semantics violates factivity. An invariantist construal of the semantics is still open, but it would naturally lead to an unorthodox invariantist view of circumstantial modals and subjunctive conditionals.
Résumé

Ce travail défend la thèse que la connaissance propositionnelle consiste en une croyance fondée sur une méthode infaillible, ou *infaillibilisme des méthodes*. Cette position appartient à la famille des analyses modales de la connaissance qui ont été développées dans les cinquante dernières années. Nous situons ces dernières dans une perspective historique large et donnons une caractérisation rigoureuse de leur forme et leurs variantes. L’infaillibilisme des méthodes est défendu comme l’analyse modale la plus simple qui satisfasse deux contraintes requises pour éviter le problème de Gettier et ce que nous appelons le problème des vérités nécessaires. Nous implémentons la thèse dans un nouveau type de modèle formel de la connaissance. Nous l’implémentons également dans une sémantique contextualiste des attributions de connaissance, mais nous laissons ouverte la question de savoir si cette dernière doit être adoptée. L’infaillibilisme des méthodes est ainsi examiné sous les angles de l’histoire de la philosophie de la connaissance, de la philosophie de la connaissance, de la logique épistémique et de la philosophie du langage, correspondant aux quatre parties de ce travail.

L’introduction précise la nature de notre projet. Nous nous intéressons à la connaissance propositionnelle, ou connaissance des faits. La connaissance des choses, ou « accointance », et le savoir-faire, ne sont concernés que dans la mesure où ils sont réductibles à la connaissance de faits, et ne sont en tout cas pas abordés directement. Nous défendons la thèse qu’une connaissance propositionnelle est une croyance fondée sur une méthode infaillible. (Précisons d’emblée que la notion de « méthode » utilisée ici, quoique dérivée d’un usage répandu par Robert Nozick, est...
spécifique à ce travail, et qu’il serait en particulier malvenu d’y associer des préconceptions cartésiennes ou ordinaires.)

Cette thèse, l’infaillibilisme des méthodes, se présente comme une analyse du concept de connaissance. Le projet d’analyse conceptuelle de la connaissance a été l’un des plus prolifiques de la philosophie de tradition dite « analytique » des cinquante dernières années, mais fait l’objet depuis peu de multiples critiques (voir notamment Williamson, 2000), tout comme les projets d’analyse conceptuelle en général. L’introduction précise dans quelle mesure la thèse défendue est ou non une « analyse » de la connaissance. Nous n’avançons pas la thèse comme une analyse du concept de connaissance qui l’expliquerait en termes de concepts mieux compris et saisis de façon indépendante de la notion de connaissance. Tout au contraire, la notion centrale de méthode que nous utilisons est essentiellement caractérisée par la thèse elle-même : l’infaillibilisme des méthodes, loin d’être une analyse du concept de connaissance, est bien plutôt une définition de la notion de méthode. Mais le fait que notre conception des méthodes soit dépendante d’une compréhension antérieure de la connaissance est compatible avec l’idée que les méthodes sont ontologiquement plus fondamentales que la connaissance, et donc que la connaissance elle-même, par opposition au concept de connaissance, se réduise à l’infaillibilité des méthodes. Pour employer une terminologie ancienne : le concept de connaissance peut être premier dans l’ordre de la connaissance alors que les méthodes sont premières dans l’ordre des choses. Notre thèse est donc une théorie, potentiellement réductive, de la connaissance, sans être une analyse du concept de connaissance.

La partie I situe l’infaillibilisme des méthodes dans la perspective de l’histoire de la philosophie de la connaissance, dont elle propose une nouvelle vue d’ensemble. Nous défendons l’idée que la connaissance était traditionnellement conçue comme une croyance portant une marque discernable et infaillible de vérité. Nous illustrons cette idée, que nous nommons l’Ur-fondationnalisme, avec la philosophie stoïcienne de la connaissance. Elle mène au final au scepticisme de Hume, puisqu’aucune de nos croyances ou presque ne porte de telle marque. A l’époque contemporaine,

La partie II défend l’infaillibilisme des méthodes, et concerne donc la philosophie de la connaissance proprement dite. Nous donnons d’abord une caractérisation rigoureuse des analyses modales de la connaissance. Nous appelons réquisit modal l’énoncé d’une condition pour la connaissance qui est remplie dans un cas si et seulement si, dans ce cas, avoir une croyance fausse ou une croyance fausse fondée sur une certaine méthode était impossible. Nous montrons que la plupart des réquisits modaux sur la connaissance développées depuis l’analyse nomologique de David M. Armstrong (1968), comme la sensibilité de Robert Nozick (1981), le fiabilisme d’Alvin Goldman (1986), ou la sûreté d’Ernest Sosa (1999b), Timothy Williamson (2000), Duncan Pritchard (2005) et Pascal Engel (2007), sont des variantes d’une forme générale dont nous identifions les traits structuraux cruciaux. Nous soutenons que si l’on doit fournir une analyse modale de la connaissance, celle-ci doit s’appuyer sur la notion de méthode, afin de ne pas être trivialisée dans le cas de croyances en des vérités nécessaires, et infaillibiliste, afin d’éviter le problème de Gettier. Nous complétons cette défense de l’infaillibilisme des méthodes en écartant des objections à l’idée que l’infaillibilité d’une méthode est nécessaire pour connaître, et à l’idée qu’elle suffit. En particulier, le problème dit de la généralité, selon lequel on ne peut proposer une condition sur la connaissance en termes de méthodes sans spécifier ces dernières indépendamment de la notion de connaissance, est écarté en identifiant les méthodes en termes de connaissance comme étant des classes
épistémiquement pertinentes de cas de croyance. Notre défense est conditionnée à la supposition qu’une théorie modale de la connaissance est souhaitable, supposition en faveur de laquelle nous fournissons quelques considérations méthodologiques.

La partie III implémente l’infaillibilisme des méthodes dans une nouvelle représentation formelle de la connaissance. Nous représentons une méthode comme une fonction des mondes possibles et d’ensembles de prémisses vers des ensembles de conclusions. Les croyances sont définies comme les conclusions atteintes sur la base d’un ensemble vide de prémisses. Nous introduisons une relation d’accessibilité qui représente la possibilité (au sens aléthique), et définissons la connaissance comme une croyance basée sur une méthode qui ne donne que des croyances vraies aux mondes accessibles. Si l’on adopte une notion « grossière » de proposition comme ensemble de mondes possibles, la sémantique qui en résulte est une extension significative de la sémantique de voisinage de Richard Montague (1968, 1974) et Dana Scott (1970). En général la sémantique valide seulement la thèse que la connaissance implique la croyance et la vérité, mais nous montrons que l’ensemble des axiomes du système modal S5 communément utilisé en logique épistémique peut être retrouvé sur la base d’une série d’idéalisations naturelles des agents. Nos modèles fournissent une perspective éclairante sur les questions de savoir pourquoi et quand les axiomes de S5 sont valides et offrent une représentation plus naturelle de notions épistémiques comme la connaissance inductive ou les cas Gettier que les modèles classiques de Jaakko Hintikka (1962).

La partie IV examine l’infaillibilisme des méthodes du point de vue de la philosophie du langage en reliant l’infaillibilisme des méthodes aux attributions ordinaires de connaissance et à leur sémantique. Nous soutenons que notre usage des verbes modaux dits « circonstanciels » (comme « il aurait pu remporter la partie ») et des conditionnels dits « contrefactuels » ou « subjonctifs » (comme « s’il avait remporté la partie, nous l’aurions félicité ») reflètent une l’existence d’une certaine notion de possibilité que nous nommons la possibilité réelle. Étant données cer-
taines connexions intuitives entre ces constructions et les attributions de connaissance, nous défendons une sémantique pour « savoir » selon laquelle la connaissance requiert d’éviter les possibilités réelles d’erreur. Les traitements communs des modaux circonstanciels et des conditionnels subjonctifs (Kratzer 2010b, Gillies 2007) suggèrent que la possibilité réelle est sensible au contexte : différents contextes de conversations peuvent compter des domaines divers de possibilités comme étant des possibilités réelles — des possibilités devant être prises au sérieux. La sémantique peut ainsi donner une version du contextualisme à propos des attributions de connaissance. Nous montrons que cette dernière évite un nombre de problèmes que rencontrent les versions du contextualisme de Keith DeRose (1995, 2009) et David Lewis (1996) — dont certains n’ont pas encore été signalés dans la littérature, comme le fait que la sémantique de Lewis viole la factivité de la connaissance. Une version invariantiste de la sémantique est néanmoins défendable, mais elle mènerait naturellement à une conception hétérodoxe des modaux circonstanciels et des conditionnels contrafactuels comme étant invariants eux aussi.


Un ensemble d’appendices complètent ce travail. Les trois premiers abordent des questions annexes et techniques liées aux parties II et III. Les deux derniers, plus longs, examinent en détail des questions plus avancées concernant les analyses modales de la connaissance, notamment
les analyses en termes de sûreté. Le premier pose un problème pour l’idée, largement partagée, que la notion de sûreté doit être conçue comme étant relative au temps : moyennant quelques suppositions naturelles, cette idée est incompatible avec la clôture épistémique, c’est-à-dire la préservation de la connaissance par la déduction. Le second examine la question de savoir si les réquisits modaux doivent être centrés sur un sujet : est-ce que pour savoir quelque chose, je dois avoir une croyance fondée sur une méthode qui ne me conduirait pas à des croyances fausses, ou qui ne conduirait personne à des croyances fausses ? La plupart des auteurs adoptent la première option sans la justifier ; nous défendons la seconde.
Chapter 1

Introduction

The central thesis of this work is that knowledge is belief based on an infallible method:

**Method infallibilism**  $S$ knows that $p$ iff $S$ believes that $p$ on the basis of a method that could only yield true beliefs.

The thesis bears on knowledge of facts as opposed to knowledge of things and knowing-how. The latter will not be discussed. They are concerned only as far as they are reducible to knowledge of facts.

1.1 **Method infallibilism and conceptual analysis**

Method infallibilism looks suspiciously like a conceptual analysis of factual knowledge. Conceptual analyses in general, and conceptual analyses of knowledge in particular, have come under severe attack in recent years. So let me clarify from the outset what a defence of method infallibilism is intended to achieve.

Method infallibilism is meant to state a truth about knowledge. Knowledge is not a concept. At first sight, it is a relation to facts or to true

---

propositions that many persons instantiate. It may turn out to consist in a family of such relations, as contextualists think, and it may turn out that hardly anybody ever instantiates it (or them), as sceptics think. But it is not a concept. According to method infallibilism, this relation (or this family of relations) is identical to the relation of believing the relevant proposition on the basis of an infallible method (or to a family of relations of that kind). The concept of knowledge, by contrast, is a type of mental entity that explains our ability to ascribe knowledge to ourselves and to others and to think and talk about it. Insofar as the concept of knowledge denotes knowledge, method infallibilism implies that the concept of knowledge denotes infallibly based belief.²

Now there are two ways in which method infallibilism could be a conceptual analysis of knowledge. On a strong version of conceptual analysis, it is assumed that our concept of knowledge is a complex mental entity, for instance an implicit theory. If method infallibilism is put forward as an analysis of this kind, then it is claimed that our concept of knowledge has concepts or representations of methods and of infallibility as its components. In the simplest case, method infallibilism is the implicit theory that constitutes the concept of knowledge. On a weak version of conceptual analysis, nothing is claimed about the structure of our ordinary concept. Rather, the analysis is put forward as a description of the referent of the concept couched in terms that are themselves better understood, or at least understood independently of the concept analysed. If method infallibilism is put forward as an analysis of that kind, it is claimed that we — philosophers, if not ordinary people — have an independent grasp of methods and of infallibility such that we can under-

². It is not trivial that the concept of knowledge denotes knowledge, as I use these terms. Epistemic revisionism is the idea that knowledge (the object of study of epistemology) is not what the ordinary concept of knowledge (the mental entity that underpins our use of "know") refers to. By analogy, revisionism about time is the idea that the ordinary concept of time does not denote time itself, the physical quantity. For instance, one may argue that the ordinary concept of time refers to a property that would be instantiated if Newtonian physics were true, but that is not instantiated in our world, though what physicists call "time" is close enough to what our ordinary concepts denotes for our ordinary time-thoughts to be true or true enough.
stand and perhaps apply the notion of an infallible method independently of our understanding and application of the notion of knowledge. We could thus test whether the two notions are coextensive by considering a range of thought experiments.

In the present work method infallibilism is not put forward as a conceptual analysis of either of these kinds. For all we will say, our concept of knowledge is atomic, and it is a primitive among our concepts. As we will point out, it is easy to think that we apply it without the mediation of other concepts. But most importantly, and most emphatically, we do not take the concept of a method to be an ordinary one, nor one that is grasped or applied independently of the concept of knowledge. (Not even by philosophers.) Rather, the concept of method is defined by its role in the account of knowledge. Thus if anything, method infallibilism defines the concept of methods rather than the concept of knowledge.

That being said, we will partly vindicate a conceptual analysis of the weak kind, namely as far as infallibility is concerned (ch. 7). If method infallibilism is right, the concept of knowledge denotes infallibly based beliefs. That does not imply that the concept of knowledge is the concept of an infallibly based belief (as the strong conceptual analysis would have it), but it implies that our ordinary applications of the concept are directly or indirectly sensitive to whether a belief is infallibly based. One way in which they are, I will argue, is through an ordinary notion of possibility of error. We have an ordinary concept of alethic possibility, and it allows us to consider possibilities of error. The intuitive appeal of some sceptical arguments shows that we are sensitive to a certain kind of alethic possibility of error when ascribing knowledge. So the infallibility component of method infallibilism is something of which we have an independent grasp.

Moreover, the history of epistemology suggests that there is, if not an ordinary conception of knowledge, at least a default theory of knowl-

---

3. See section 4.3.2. Of course we need to have some mental representations of people, of our surroundings and of objects of knowledge in order to apply the concept. But no specific representation is needed beyond that of persons; in particular, no representations of mental states or properties such as beliefs or methods.
edge that has repeatedly been endorsed by philosophers in the Western
tradition (ch. 2). The default theory, which I call Ur-foundationalism, is
that knowledge requires a discernible and infallible mark of truth. The
discernibility requirement, when applied consistently, can at most be
satisfied by internal states. The view is thus naturally led to a form of in-
fallibilist internalism according to which one knows only if one’s internal
mental state is an infallible mark of truth. (This in turns warrants scep-
ticism.) Now if we extrapolate from default (Western) epistemological
theories to the ordinary conception of knowledge, we should expect that
if there is anything like an ordinary notion of method, it is going to be an
internalist one. But we should certainly not tie ourselves to such a notion
from the outset. So, again, the notion of method is to be understood as
a purely theoretical one.

1.2 What methods are

Methods are epistemically relevant types of belief cases. Our start-
ing point is the idea that knowledge is a matter of impossibility of error:
whether one knows in a case depends on whether one believes truly in
a range of possible variants of the case. Call these variants the relevant
alternatives to a case. Relying on our intuitive judgements about knowl-
dge, we get an idea of which alternatives are relevant and which are not.
Namely, relevant alternatives to a target case are cases of belief whose
history resembles that of the target belief in a number of respects. The
relevant respects of history involve aspects of the environment in which
the belief is formed and aspects of the cognitive processes that lead to
its formation or support it. But it is doubtful that the relevant respect

4. That said, I do not think that there is in fact an ordinary notion of method, in-
ternalist or otherwise. Rather, we may have an internalist heuristic when ascribing
knowledge, to the effect that if two subjects are internally alike, and one is in error, we
deny that the other knows. (I myself endorse a weaker version of the heuristic in sec.
4.3.4.)
5. A case is defined as a centred world with a proposition singled out (sec. 3.1).
6. Relevant alternatives are alternative cases (as in Goldman, 1976, 774–8), not alter-
 natives to the proposition believed.
of similarity can be factored into a conjunction of similarity under various respects, such as similarity in the environment and similarity in the inner cognitive processes. Rather, we should simply say that relevant alternatives are the ones in which a belief has one of the same relations to its external and cognitive environment as the target belief — where the relation belongs to the kind of relations of a belief to its external and cognitive environments that are relevant to whether one knows. Relations of this kind are what we call methods.

At one extreme, it could in principle turn out that the only relevant type for all knowledge cases is that of knowing. That is, there is a single method that all and only cases of knowledge share. (Non-knowledge cases could also share a unique method, or not.) But that is implausible. Consider a series of cases in which one is asked to visually identify an object at various distances. We may assume that the person knows when the object is close enough, but fails to know when the object is far away. It is natural to take the series of cases as instantiating a range of relations to the environment that resemble each other more than the first ones resemble cases of mathematical proof, despite the fact that the first ones and the mathematical proof cases are both knowledge cases. It is plausible that there are epistemically relevant types of relations to the environment that are more specific than the relation of knowing itself.

In ordinary language, the closest we get to denoting methods or method types is by expressions entailing knowledge such as calculating a sum, remembering distinctly an episode of a recent dinner, identifying a familiar object by sight at close distance, and by expressions entailing failure of knowledge such as being deceived about the results of a meeting, guessing what an object is, having a dim memory of a party. But again, we should not rely on common sense types to figure out what methods are.

Given that methods are types of beliefs, there is no need to assume that each belief is based on one method only. A belief is an instance of

---

7. See sec. 4.1.3.
8. Accordingly, methods are modelled as functions from world and beliefs to beliefs (ch. 5).
knowledge iff there is at least one method on which it is based that is infallible — that is, iff it belongs to at least one relevant type all instances of which are true beliefs. It may at the same time belong to fallible types. For instance, we may think that one of the relevant types is that of belief itself. Each belief is based on the method of believing, but no belief is knowledge in virtue of being based on it. Typically (but not necessarily), the relevant types of belief in virtue of which a given belief is knowledge will be among its narrowest ones.

1.3 Methods infallibilism and the theory of knowledge

We introduce and constrain the notion of methods in two ways. First, through method infallibilism itself, the notion is tied to our intuitions about knowledge and to a notion of possibility (ch. 4). Second, we give a formal representation of methods which tells us what structure something must have if it is to correspond to the notion (ch. 5). In particular, we introduce structural relations among methods themselves which go beyond what method infallibilism requires. Moreover, the formal representation allows us to individuate a few specific methods.9

None of this guarantees that there are methods, that is, that there are a range of natural properties of belief that correspond to our notion. What makes a property a natural one is that it enters in simple relations with other properties. So the hope is that the notion of method can be put to use in a variety of contexts beyond method infallibilism. The formal models we put forward are a first step in that direction: they set out a number of systematic relations among methods themselves and explain some important links between belief and knowledge (ch. 5). A second step is our hypothesis that the infallibility of methods is assessed with respect to a space of real possibilities that is independently involved in

9. The structural relations in question are method union and composition (see sec. 5.2.2). The methods we individuate are Pure Reason, Deduction, and Introspection and Confidence methods (see sec. 5.5).
our counterfactual reasoning (ch. 7). Further steps to vindicate methods would relate them other notions than knowledge, belief and real possibility. They may be linked to action, through pairing epistemic methods with decision-making methods to explain action. They may be related to reasoning through diachronic belief-formation and belief-revision methods. They may be related to norms of belief, if we explain why beliefs ought only to be based on infallible methods. They may be related to value or credit, through an account of why true beliefs based on such methods are distinctively valuable or in some sense creditable to one. I will not explore these questions in the present thesis, but they will have to be investigated in future work.

The exuberant profusion of theories of knowledge sparked by the Gettier problem has made it clear that one should not put forward an account of knowledge simply on the basis of its ability to yield intuitive judgements about a few problematic cases. Accounts of knowledge should also be motivated; they should make sense of the fact that knowledge is a distinguished property and of the role we expect it to play in thought and action. Call this (admittedly vague) desideratum the Craig desideratum.\(^{10}\) The present work meets it in a minimal way by defending method infallibilism not on the basis of a few problematic cases, but rather on the basis of general constraints on the structure of accounts of knowledge in terms of possibility of error. The desideratum would be fully met if we could put method infallibility to use in the explanation of action, the norms of belief, assertion and practical reasoning, or in an explanation of the value of knowledge. This will not be done in the present work, but we hope that the reader familiar with recent epistemology will have a feeling of how these developments can be pursued.

To sum up, even though the notion of method is defined through method infallibilism itself, this does not make methods infallibilism trivially true, nor does it make it untestable. Method infallibilism will turn

---

10. It was notably put forward by Craig (1990), 1-2. Mark Kaplan (1985) advanced a similar idea, though in order to argue that knowledge was not an interesting property since it failed to satisfy the desideratum.
out false if there is no range of sufficiently natural properties of beliefs that validates it. Method infallibilism cannot be tested by comparing our intuitions about knowledge with our intuitions about methods (since we have none of the latter), but it is tested by whether the notion of method integrates smoothly with other notions like belief and action.

Now if the project succeeds, this means that there are methods, and that knowledge is belief based on infallible methods. If so, it is open to consider that method infallibilism provides a reductive account of knowledge, in the sense that methods and impossibility would be ontologically prior to knowledge. This could be the case even though the theoretical concept of method is ultimately derived from our ordinary concept of knowledge. To use an older terminology: though knowledge is the first thing in the order of discovery, methods may be the first thing in the order of things. I am not going to defend the ontological priority thesis in the present work. But if it was true, it would give a sense in which methods infallibilism is an analysis of knowledge, as opposed to an analysis of the concept of knowledge.

1.4 Overview

This work is divided into four parts: history, epistemology, formal epistemology and semantics. The first part situates infallibilism in a broader historical perspective. The second part defends method infallibilism on epistemological grounds. The third part implements method infallibilism in a new type of model for belief and knowledge. In the fourth part, ordinary knowledge ascriptions are linked to a notion of possibility that transpires from the semantics of modals and subjunctive conditionals. Each part can be read fairly independently, though their present order is the most natural one.

Three appendices (A, B, C) provide supplementary materials for chapters 3, 4 and 5, respectively. Two further appendices (D and E) are stand-alone treatments of advanced questions on safety-style conditions on knowledge, of which method infallibilism is an instance.
1.4. Overview

Chapter 2 provides an alternative to a widespread view of history of epistemology according to which Edmund Gettier’s (1963) paper refuted the traditional analysis of knowledge. On the view we put forward, Gettier did refute the justified true belief analysis of knowledge, but the analysis was not the traditional one. Rather, on the traditional conception, knowledge consists in having a discernible and infallible mark of truth. Virtually all philosophers endorsed a variant of the view until Hume and many beyond. The discernibility requirement leads to internalism and, in conjunction with the infallibility requirement, to scepticism. The story explains why contemporary epistemology is divided into two trends: internalists drop infallibility in order to preserve discernibility, and externalists drop discernibility to preserve infallibility. Fallibilist internalism is the kind of view that runs into Gettier problems. The historical perspective thus provides a motivation for an infallibilist view of knowledge. It may also give some insight into the ordinary conception of knowledge, insofar as the recurrence of the discernible and infallible mark view is an indicator of its intuitive appeal.

Chapters 3 and 4 provide a defence of method infallibilism. Method infallibilism is a species of modal requirement on knowledge, that is, a requirement that is satisfied in a case iff error is avoided over some relevant range of alternative cases. Chapter 3 characterises such requirements in rigorous terms, and gives some methodological considerations in favour of trying to provide an account of knowledge in modal terms. We show that most major modal accounts of the recent literature (relevant alternatives, reliabilism, sensitivity, safety, and so on) share a common form. The review allows us to point out important structural properties of such requirements. Three are worth mentioning here. Proposition-centred requirements are such that a belief that \( p \) satisfies them as long as false belief in the same proposition is avoided at relevant alternative cases. Such requirements run into what I call the necessary truth problem — see the next paragraph. Subject-centred requirements are requirements that a belief that \( p \) satisfies if false belief by the same subject is avoided at relevant alternative cases. (The question whether subject-centring is justified is
Variable requirements are those for which the range of relevant alternative cases depends on the particular proposition believed — a paradigmatic example is Nozick’s sensitivity. Variable requirements typically violate the closure of knowledge under competent deduction. In appendix D, we argue that under natural assumptions, if the notion of “close” possibility used in safety requirements is conceived as being time-relative, safety is a variable requirement and consequently violates deductive closure. Appendix A provides details of the various semantics of subjunctive conditionals used in sensitivity and safety conditions.

The case for method infallibilism is made in chapter 4. Assuming that a modal account of knowledge is to be given, we first argue that all proposition-centred requirements face the necessary truth problem: they are trivially satisfied by beliefs in necessary truths. The problem is well known but its consequences underestimated. Secondly, we argue that reliability requirements — requirements that fall short of infallibility — cannot avoid Gettier problems. Thus method infallibility is necessary for knowledge. Third, we argue that a sufficient condition for knowledge should be simple: conjunctions of conditions are typically liable to coincidental satisfaction, which provide counter-examples to their sufficiency. The simplicity idea pleads against adding any further condition on method infallibilism; if method infallibilism was not sufficient, one would rather have to replace method infallibility altogether with a stronger requirement. I discuss and reject major objections to the sufficiency of method infallibilism, and tentatively conclude that is is sufficient for knowledge as well. The overall argument is conditional and abductive. It is conditional on the idea that knowledge should be accounted for in modal terms; some methodological reasons to think so are given in ch. 3, but they are less than compelling. It is abductive as far as sufficiency is concerned because it is merely confirmed by the failure of salient objections to it. Appendix B explains why I find the “warrant entails truth” formulation of infallibilism insatisfactory.

Chapter 5 introduces models of knowledge and belief based on the
method infallibilist idea. We give a formal representation of methods and of their infallibility, and we define knowledge and belief in those terms. We argue that epistemological notions and problems like Gettier, inductive knowledge, fallible justification, or failure of logical omniscience are represented in more satisfactory ways in methods models than in standard Hintikka-style models for knowledge. The resulting models are a significant extension of the so-called “neighbourhood models” for modal logic introduced by Richard Montague and Dana Scott. In general, the models only validate the claim that knowledge requires true belief; but we show that a full S5 system can be derived from a set of natural idealisations of the agent’s methods. The derivation gives an insight into why and when the S5 axioms should hold of knowledge, and a vindication of their widespread use in formal representations of knowledge. We additionally derive a number of intuitive connexions between belief and knowledge, as well as some properties of knowledge that can only be stated by reference to methods, such as closure of knowledge under competent deduction as distinct from logical omniscience. Appendix C provides further developments that could be omitted from the main text.

Chapters 6 and 7 relate method infallibilism with the semantics of “know”. As said in this introduction and as argued in chapter 4, I do not think that there is an ordinary conception of methods. But in these chapters, I will argue that there is an ordinary notion of possibility — which I call real possibility — that is manifested in our use of so-called “circumstantial” modals and of subjunctive conditionals. Our patterns of knowledge attribution, and in particular the intuitive appeal of sceptical arguments, show that we are sensitive to real possibility of error when ascribing knowledge. The real possibility account is defended in contrast to recent modal contextualist accounts of knowledge ascriptions. Chapter 6 presents DeRose’s and Lewis’s accounts. Both implement the modal idea that knowledge requires avoidance of error at a relevant set of alternative cases in a contextualist semantics of “know”. We point out a number of difficulties for them, and we stress that each is led to posit a context-sensitive semantic parameter that is specific to “know”. By
contrast, on the real possibility account presented in ch. 7, ascriptions of
knowledge rely on an independently provided set of possibilities. However, invariantism about real possibility is also an option, and can be used
to defend invariantism about knowledge attributions. Here I will only
defend the claim that one should endorse contextualism about “know” if
contextualism about claims of real possibility is true.

Appendices D and E treat advanced topics on safety-style conditions,
including method infallibilism. Appendix D argues that under natural
assumptions, a time-relative conception of safety violates deductive clo-
sure, contrary to what many assume. Appendix E argues against the
common practice of using subject-centred modal requirements.

---

A detailed table of contents is given at the end of the volume. Contents are also
summarised in the conclusion. A number of symbols are introduced in chapters
3 and 5. They are collected in a list of symbols, provided along the lists of tables
and figures.
Part I

History of Epistemology
Chapter 2

The Legend of the Justified True Belief Analysis

In the last few decades, a certain picture of the history of epistemology has gained wide currency among epistemologists. The Legend, as I will call it, is summarised in the claim that:

Edmund Gettier’s landmark paper successfully refuted the traditional analysis of knowledge.\(^1\)

As my label indicates, I think that the Legend is false. Not that Edmund Gettier did not refute the justified true belief analysis of knowledge. He did, but the analysis was not the traditional one.

While the Legend figures in about any contemporary handbook of epistemology, I do not, in fact, expect a strong resistance to the claim that it is false. The Legend is widespread not because it has been powerfully defended — to my knowledge, it has never been defended at all — but rather because no better picture has been available.\(^2\) Such a picture is precisely what I intend to offer here. Call it the New Story.

Like the Legend, the New Story is painted in broad strokes. It ignores a lot of historical detail and involves a significant amount of rational

---

2. Thus Mark Kaplan’s (1985, 352-353) criticisms of the Legend have been largely ignored.
reconstruction. Yet I think it offers a recognisable portrait of the history of epistemology. It also provides an illuminating perspective on the present of the discipline.

The outline of the New Story is as follows. There is a traditional conception of knowledge, but it is not the justified true belief analysis Gettier attacked. On the traditional view, knowledge consists in having a belief that bears a discernible and infallible mark of truth. Roughly, a mark is discernible if a sufficiently attentive subject believes that a belief bears it if and only if it does, and it is infallible if a belief that bears it cannot be false. Hellenistic philosophers called such a mark a criterion of truth. Since a belief that bears a mark of that kind cannot be false, the truth condition on knowledge is superfluous. From Aristotle to Russell, many philosophers have endorsed a version of the view. My hypothesis is that before Hume, if not before the XXth century, all or almost all have. Since the conception appears to be the one philosophy started with, and because it is foundationalist — as we will see —, I call it Ur-foundationalism.

Sceptics have put the view under pressure since its inception. Not that they rejected it; rather, they made it difficult to accept it alongside any other positive doctrine. But we owe it more particularly to Hume to have shown the extent to which we lacked discernible and infallible marks of truth. For while ancient sceptics focused on putative marks of truth for perceptual beliefs and for ethical and philosophical positions, Hume’s problem of induction showed that even if perceptual knowledge of our present surroundings was granted, very few of our putative knowledge could be recovered. Scepticism aside, Hume’s realisation prompted three reactions.

The first was idealism and its more modern cousin, verificationism. Such views hope to deny Hume’s predicament. If the best part of our beliefs turn out to be about our own ideas or observations, we may have discernible and infallible marks of truth after all. For, it is thought, a sufficiently attentive subject can discern the presence of a given idea or observation, and that presence would at the same time ensure the truth of the corresponding belief. These views were strikingly popular from the
end of the XVIII\textsuperscript{th} century to the begin of the XX\textsuperscript{th} (and beyond). But by the second half XX\textsuperscript{th} century, philosophers in the analytic tradition widely took them to fail both on their semantic and epistemological sides.

The second was pure internalism. A non-sceptic who wants to maintain that knowledge requires a discernible mark of truth cannot require it to be also infallible. The idea that beliefs could be fallibly justified was held by various philosophers throughout history. In the second half of the XIX\textsuperscript{th} century, it was prominent in the works of Peirce, Meinong, and later in those of Popper. But these philosophers did not claim that fallibly justified beliefs were knowledge. By the mid-XX\textsuperscript{th}, Malcolm, Ayer and Chisholm went further and took fallible justification to be sufficient for knowledge. Of course, truth is needed as well, so truth appeared as an independent condition. That is the familiar justified true belief analysis that Gettier refuted.

The third was externalism. To avoid the Gettier problem, one has to maintain that knowledge requires an infallible mark of truth. But to avoid scepticism as well, one has to give up the idea that the mark should be discernible. After the publication of Gettier’s paper, several philosophers put forward conditions on knowledge that were infallible but indiscernible, in the sense that even an ideally attentive subject could mistakenly believe that they obtain while they do not. Peter Unger first stated the idea in most general terms: knowledge requires non-accidentally true belief. Whatever fact in virtue of which a belief is non-accidentally true is an infallible mark of its truth. But typically, such a fact will not be discernible, even to attentive subjects. The infallibilist conditions on knowledge that appeared in the contemporary writings of Armstrong, Goldman, or Dretske were indiscernible facts.

These are the three main options. But a final twist has to be mentioned. One could maintain that knowledge requires both a discernible mark and an infallible one, while giving up the idea that a single fact plays both roles. So there would be two distinct conditions on knowledge, an “internal” and an “external” one. A number of so-called “fourth-clause” theories that were developed in the wake of the Gettier problem
are instances of this strategy. I call them mixed internalist views. Such views face what we may call the secondary Gettier problem. If the two conditions are genuinely independent, they can be both satisfied by mere coincidence. In such cases, the subject does not know. So the conjunction of conditions is not sufficient. I will suggest that mixed internalist views cannot avoid the problem without falling back into either Ur-foundationalism or externalism.

To sum up, the New Story divides post-Humean epistemology into three groups. Ur-foundationalists take knowledge to require an infallible and discernible mark of truth. They are either sceptics or idealists/verificationists. The two other groups give up some aspect of the traditional view. Internalists maintain the discernibility requirement. They either give up the infallibilist requirement or introduce it as an independent condition. Externalists maintain infallibility but acknowledge indiscernibility.

An extended defence of the New Story is well beyond my abilities and the scope of this work. I will be content to set it out as a hypothesis worthy of further testing and elaboration. The best way to do this, I think, is to make its theoretical structure as clear as possible. I will thus present idealised versions of the positions before illustrating them with representative historical cases. A few salient objections and counterexamples will also be discussed. Overall, I will favour simplicity and clarity over accuracy. Though I will generally avoid cumbersome hedging, I am well aware that the ascription of an idealised view to any particular philosopher — not to mention classes of philosophers — is bound to face a number of opposite statements that have to be ironed out in some way or other. The understanding of history I am aiming at here is only the kind one gets from a simple model whose limits (or outright inadequacy) are easy to see.

---

3. There are two reasons for the label. First, it roughly fits use: since nowadays nobody is a pure internalist as defined (that is, a defender of the justified true belief account), such views are what epistemologists have in mind when (and if) they talk about “internalism about knowledge”. Second, I think that like pure internalism, these views face the Gettier problem, albeit in a more complex form.
2.1. The Legend and its puzzles

I will proceed as follows. In section 2.1, I point out a number of facts which would be puzzling if the Legend was true. Section 2.2 introduces Ur-foundationalism, which I take to be the traditional conception of knowledge. It is illustrated by the Stoic conception of the criterion of truth in section 2.3. Section 2.4 presents Hume’s challenge to Ur-foundationalist views and the idealist reaction to it. Section 2.5 retraces the origins of the Justified True Belief view that Gettier’s paper refuted, and provides a perspective on the contemporary divide between internalism and externalism. As will become clear by sections 2.4–2.5, the New Story’s characterisation of positions around questions of (in)discernibility and (in)fallibility leads to some reconception of the layout of the contemporary landscape. It also indicates promising avenues and likely impasses. I conclude in section 2.6 by drawing a few lessons the New Story would teach us, if true.

2.1 The Legend and its puzzles

The Legend is the claim that “Edmund Gettier’s landmark paper successfully refuted the traditional analysis of knowledge.” 4 The claim can be found in about every contemporary handbook of epistemology. 5 It has two components: (1) the traditional conception of knowledge was the justified true belief analysis, and (2) Gettier refuted the justified true belief analysis of knowledge. As I have said, I think the Legend is false. More precisely, I think that one cannot hold both (1) and (2). Let me first clarify this point.

My view is roughly that on the traditional conception, knowledge requires something like an infallible justification. But if that is so, could we not maintain that the justified true belief conception is the traditional

5. For a recent sample, see Williams (2001, 16), Moser (2002, 4, 29), Pritchard (2006, 25), Pritchard and Neta (2008, 5-6). The present author is not an exception (Dutant and Engel, 2005, 36). Williams (26n) and Pritchard, as well as Dancy (1984, 22) avoid a straightforward endorsement. A notable exception is Steup (2006), who avoids calling the justified true belief analysis “traditional” entirely.
one, in which justification was construed as infallibilist initially and as fallibilist more recently? We could, and some proponents of the Legend do. But then we could not say that Gettier refuted it. For the Gettier problem does not arise for infallibilist views, as we will see (sec. 2.5). We would have to say instead that Gettier refuted the fallibilist version of the traditional analysis. This would be misleading, since the fallibilist version of the traditional analysis is not itself “traditional” at all. And this would fail to cut epistemology at the joints, since it would lump together accounts that face very distinct problems: the traditional infallibilist one, which faces sceptical challenges, and the modern fallibilist one, which faces Gettier counterexamples. That is why I prefer to reserve the label “justified true belief” for modern fallibilist accounts. (On infallibilist accounts, the truth condition is redundant, so “infallibly justified belief” would be a more appropriate label for them.) On this use, we can say that Gettier refuted the justified true belief analysis, but not that the justified true belief analysis was the traditional one. Whichever terminology one adopts, either (1) or (2) fails, so the Legend is false. Or so I will argue.

My main argument against the Legend is simply that the New Story is true. That is the topic of the next sections. In the present one I point out some facts that should puzzle anyone who believes the Legend to be true (2.1.1 to 2.1.4). I will also sketch a history of how the Legend appeared (2.1.5).

2.1.1 Why is it so hard to find any clear statement of the justified true belief analysis before the XXth century?

First, one hardly finds any clear statement of a justified true belief analysis prior to Ayer’s and Chisholm’s. Plato, Kant and Russell could
be cited. But I do not know any such statement from another major philosopher, and I suspect that if there was, it would have been widely reported by now. I also doubt that any will be found in any philosophical handbook or dictionary prior to 1950. So even if we conceded the cases of Plato, Kant and Russell, that would be strikingly few. What about Aristotle, Epicurus, the Stoics, the Sceptics, Thomas, Occam, Scotus, Descartes, Spinoza, Leibniz, Locke, Berkeley, Hume, or Reid, to name a few? Philosophers are not normally shy to state the obvious. Their silence is thus surprising even if they took the analysis to be trivially true. One would particularly expect it to have cropped up in the discussion of sceptical arguments, given that a fallibilist construal of justification would provide a quick reply to sceptics. But it did not.

As for Plato, Kant, and Russell, each attribution is debatable. I need not discuss the cases of Kant and Russell in detail here. Both were post-Humean epistemologists. The core idea of the New Story is that the justified true belief analysis only arose as a reaction to Hume’s predicament. In my view, the first to have adopted it are Malcolm, Ayer and Chisholm. If Kant and Russell turn out to have predated them, most of the New Story remains in place. But let us examine the case of Plato.

Plato famously put forward the definition of knowledge as “true opinion with an account”, where an “account” (logos) is some sort of explanation of the thing known. It is unclear whether he endorsed it. In the Theaetetus (201-210), the definition is found unsatisfactory, but in other places Plato appears to defend it. So it is open to argue that what is unsatisfactory in the Theaetetus is not the definition itself but rather the three proposed accounts of “account”.

7. Thus Newman (2007, 319–321) finds it surprising that Locke appears not to have endorsed the “traditional” justified true belief view. (Locke’s stated definition is: “knowledge is the perception of the agreement or disagreement of two ideas,” Essay, IV, 1, §2.) Newman removes the supposed anomaly by arguing that Locke did endorse the justified true belief view. But he makes clear that Locke’s “justification” is infallibilist, so it is not a justified true belief analysis in the present sense.

8. See Meno 98a2, Phaedo 76b5-6, Phaedo 97d-99d2, Symposium 202a5-9, Republic 534b3-7, Timaeus 51b6-e6.

9. See Chappell (2009, sec 8). Another possibility is that the definition was defended.
Whether Plato endorsed it or not, it is debated whether the definition corresponds to the contemporary justified true belief analysis of propositional knowledge. A first issue is whether the topic of Plato’s discussion is knowledge in general or rather what we would call “expert knowledge” or “science” — the examples of “knowledge” given at the beginning of the dialogue are geometry, astronomy, harmonics, arithmetics, and shoemaking (145d1, 146c-d). If the dialogue bears only on expert knowledge, the “account” requirement may be specific to it. A second problem is whether Plato has in mind propositional knowledge, objectual knowledge, or both: Theaetetus, the sun and virtue are among the examples he gives of things known. If Plato’s definition does not cover propositional knowledge, it does not correspond to the modern one. But even if Plato’s definition applies to ordinary propositional knowledge, a crucial question remains: would Plato endorse a notion of “account” on which a belief with an account can be false? If not, his view is infallibilist and is not of the kind that the Gettier problem affects. 

by some contemporaries of Plato, since it is introduced as “something [Theaetetus] once heard someone saying”, 210c8. Burnyeat (1990, 164–173) rejects its attribution to Antisthenes, the most likely known candidate.

10. Burnyeat (1970), Annas (1982), and Nehamas (1984) defend the idea that the Theaetetus is about expert or scientific knowledge. Burnyeat (1990, 216-218) suggests that we should regard Plato’s theory as a theory of (what we would call) understanding rather than of (what we call) knowledge; see also his Burnyeat (1981), and see Barnes (1980, 205-206) for contrary advice. Fine (1990, 107) agrees that the sort of account Plato has in mind is an explanation but argues that this does not prevent the assimilation of his definition with the justified true belief analysis: “His view is rather that justification typically consists in, or at least requires, explanation. For Plato, I am justified in believing $p$ only if I can explain why $p$ is so; I am typically justified in claiming to know some object only if I can explain its nature or essence.” See also Vlastos (1985) and Woodruff (1990, 60–67) on expert knowledge in Plato’s early dialogues.

11. White (1976, 176ff) and Nehamas (1984) object to comparisons with the modern account on this basis. Fine (1979, 366–7) argues that Plato uses “$S$ knows $x$” and “$S$ knows what $x$ is” interchangeably, and that the latter expresses propositional knowledge. However, she concedes that Plato is not interested in just any claim to know that $p$, but in explanatory knowledge of the nature of entities one claims to know. But on her view, Plato gives a justified true belief analysis of this subclass of knowledge.

12. If an “account” is an answer to the question why, there is a natural line of thought that leads to the idea that false beliefs do not have an “account”: namely, a question “why $p$?” presupposes that $p$ is true, and is without answer if $p$ is false, just has “who ate the cookie?” has no answer if the cookie was not eaten. Thanks to John Hawthorne
2.1. The Legend and its puzzles

Was Plato an infallibilist? Knowledge is said to be “free from error” in the *Theaetetus* (152c5-6) and “infallible” or “not in error” in the *Republic* (477e6-7). But these may simply restate the fact that what is known is true.\(^{13}\) Another line of argument would be to consider the dialectic that leads to the “true opinion with an account” in the *Theaetetus*. At 201a-c, Socrates quickly dismisses the idea that knowledge is simply true opinion with the famous example of the Jury: a jury that is persuaded to hold a true opinion does not know the truth. The dialogue then moves on to consider the tripartite definition. This suggests that what a jury is offered in a trial is not an “account”. For if it was, the proposal would already have been refuted. Now Socrates says that what the jury receives is persuasion, not teaching. So persuasion would not provide an “account”. Why not? Socrates only says that “[the litigants and speech makers] persuade others by means of their art, not teaching them, but making them judge whatever they want to judge” (201a9-10). One thought here would be: if the litigants make the jury judge whatever they want them to judge, they may make them judge something false. That is why it is not an “account” in the relevant sense; the kind of “account” considered in the tripartite definition should be such that a belief with an account cannot be false. Attractive as it is, that line of argument is not without problems. Socrates’ idea might just as well be that litigant’s defences are not even fallibly reliable. The text only makes clear that Socrates takes the jury’s opinion not to be adequately founded; we are left to guess what an adequate foundation would require.\(^{14}\) Moreover, the ensuing discussion of the three proposed accounts of “account” does not display any interest in the question whether having an “account” guarantees truth. So it is

\(^{13}\) Vlastos (1985, 12-13) defends the infallibilist reading, Fine (1990, 90) rejects it. Though Fine (1990, 114), defending a coherentist interpretation of Plato, writes that “[on his view] knowledge requires, not a vision, and not some special sort of certainty or infallibility, but sufficiently rich, mutually supporting, explanatory accounts”, she does not raise the question whether a belief can be part of such a coherent net of accounts and yet be false.

\(^{14}\) See Burnyeat (1980, 177-8; 1990, 124-125).
dubious that the question is on Plato’s mind in the Jury case.  

There is no clear-cut textual evidence that Plato endorsed an infallibilist conception of knowledge. But we may find indirect evidence that he did not endorse a fallibilist one in the fact that fallibilism simply did not seem to be a live option for his immediate successors. For, as we will see, Aristotle and the main Hellenistic schools all endorsed infallibilist views. That would be surprising if Plato had presented a fallibilist option. But it may be that Plato simply did not raise the question whether a belief with an “account” may be false, and thus did not clearly endorse an infallibilist view either.

2.1.2 Why did nobody notice Gettier cases?

Second, nobody noticed Gettier cases before the XXth century. The Theaetetus was one of the most widely read philosophical texts throughout history. It ends with the suggestion that having a true belief with a “account” is not sufficient for knowledge. That is, if the Legend is true, it points right in the direction of Gettier cases. Still, nobody thought of them before Meinong, Russell and Gettier did.

One should certainly not underestimate the difficulty involved in philosophical imagination. Nowadays it may seem obvious that one’s experience could be just the same and the external world not exist or that the future could be radically different from the present. But that was not so before Descartes and Hume spelt out thought experiments that convincingly established such possibilities. And it was no small feat of imagination to do this. Similarly, one should not retrospectively assume that Gettier cases were easy to invent.

---

15. A different account of the passage stresses Socrates’ claim that the jury is persuaded “about things which it’s possible to know only if one has seen them and not otherwise” (210c). Burnyeat (1980, 178-179; 1990, 125) reads this as implying that no testimony can provide knowledge, which would be in conflict with Plato’s apparent admission that teaching can transmit knowledge. Barnes (1980) rejects the implication. Why would testimony fail to transmit knowledge? Because it fails to provide a kind of understanding appropriate to things known, says Burnyeat (1980, 186-7). But another proposal would be that if fails to give knowledge because it fails to guarantee truth.
Still, it is surprising that nobody noticed them, if the Legend is true. For Gettier cases are not *outlandish*. We encounter some fairly regularly. Testimony gives many occasions to find oneself in a Gettier-style situation. My brother tells me that there are fruits in a tree near the lake; there are indeed, but it later turns out that the tree he had in mind was not the one where the fruits were. Thus many a time a false belief grounds truthful testimony. Trials are also a source of Gettier-style situations. A jury may clear a defendant on the basis of the apparently reliable testimony of a friend; the testimony turns out to be false, but the defendant is innocent none the less. In fact, the trial example almost literally appeared in the *Theaetetus*, as we have seen. So why did nobody notice?

If the Legend is true, one wonders how the justified true belief analysis has been able to survive up to the XXth century. Not only did Plato raise the question whether the definition was sufficient, but he gave a counterexample, and real-life ones are regularly to be found.

### 2.1.3 Why were Gettier-style cases presented as cases of true belief that is not knowledge?

Third, some recognisably Gettier-style cases were pointed out before Gettier’s paper. But they were pointed out as cases of *true belief* that is not knowledge, not as cases of *justified* true belief.

The first is Plato’s Jury: an able lawyer persuades a jury to reach a verdict which is in fact the right one (*Theaetetus* 201c). As Burnyeat (1980, 177-178) insists, nothing in Plato’s description indicates that the Jury’s decision is irrational; rather the opposite. So the case can be seen as one of rational but accidentally true belief.

The second and third are Meinong’s cancelling errors case and auditory hallucination case:

Truth may not be accidental, thus not comparable to a calculation the result of which is correct as a result of mistakes which accidentally cancel each other out. An a posteriori counterpart to this a priori example: somebody who has a ringing in his
ears thinks because of this that he hears a house-bell ringing and is right because someone is accidentally actually ringing a bell.  

(\textit{Meinong, 1973, 619})  

Meinong also gave an analogous case involving an Aeolian harp: someone who is accustomed to an Aeolian harp has become hard of hearing and has auditory hallucinations while the harp is actually making sounds (\textit{Meinong, 1906, 30–31; 1973, 398–399; the cases were pointed out by Chisholm, 1977, 104}).

The third and fourth are Russell’s. One is the Balfour/Bannerman case. A man believes rightly that the last name of the Prime Minister in 1906 began with a B, but he believes it because he thinks that Balfour was Prime Minister at the time, whereas it was in fact Campbell-Bannerman (\textit{Russell, 1912, ch. 13; 1948, 170–1}). The other is the Stopped Clock. A man happens to look at a stopped clock right at the time which the clock indicates (\textit{Russell, 1948, 170–1}). Russell does not specify whether the first has a good reason to think that it was Balfour, and whether the second is justified in trusting the clock. But he does not imply the contrary.

Now the interesting point is that Russell and Plato put forward these cases as cases of \textit{true belief that is not knowledge}, not of \textit{justified} true belief. \footnote{As for Meinong, he distinguished “evidence” from “conjectural evidence”. The former is infallible, the second is not. Judgements about the external world are conjecturally evident at best, and they are not knowledge proper (Meinong, 1973, 458-9). His views would have entailed that the subjects in his examples have a belief which is true and \textit{conjecturally evident}, though he does not explicitly describes them as such. The fact that they have conjectural evidence only would have been sufficient for him to explain why the subjects do not know. (Thanks to Kevin Mulligan and Fabrice Teroni here.)} And both go on to present what is allegedly the justified true belief analysis \textit{right after they have presented the cases} (\textit{Theaetetus 201, Russell, 1948, 171}). If the Legend is true, both introduce the analysis just after having presented counterexamples to it.

\footnote{16. Thanks to Kevin Mulligan for the translation, reference and a discussion of Meinong’s view on the case.}
2.1.4 Why was the justified true belief analysis not introduced as the traditional one?

Fourth, when Ayer and Chisholm presented the justified true belief analysis, they did not present it as the traditional one. Of course they may have been unwittingly reinventing the wheel. But we should at least consider whether we are not the ones who are mistaken in seeing their analysis as the traditional one.

Another view of “the traditional conception”: knowledge as a *sui generis* mental state

It was certainly not common knowledge at the time that the justified true belief analysis was the “traditional” one. Here is what we can read in an introduction to epistemology published by the Oxford philosopher A. D. Woozley in 1949:

> The notion that our cognitive activities can be sharply distinguished into kinds which are fundamentally different from each other, knowing on the one hand and believing on the other, has a long philosophical history, and has endured a less chequered career than most philosophical notions of its antiquity. According to the traditional view, which derives from Plato, knowledge and belief are mental faculties, each *sui generis*, no more to be defined one in terms of the other than are, say, love and friendship. (Woozley, 1949, 176)

Woozley thus ascribes to the tradition a view we would more readily associate with his predecessors John Cook Wilson (1926, 34-47), Harold A. Prichard (1950, 86) and with Timothy Williamson (2000), according to which knowledge is a *sui generis* mental state that cannot be defined in terms of belief, let alone as justified true belief. Since knowledge requires truth, if knowledge is a *sui generis* mental state, it is an *infallible* mental state. Prichard himself (1950, 86-7) ascribed the view to Plato.

---

18. See also Price (1934, 229–31).
(Not unreasonably so: many passages in the Republic appear to claim that knowledge and belief do not even have the same kind of objects.\(^{19}\)) So it was far from obvious by the mid-XX\(^{th}\) century that the justified true belief analysis was the traditional one.

**Ayer on the tradition and his view**

Ayer’s Problem of Knowledge opens with an attack on the infallible mental state view (Ayer, 1956, 15-23). Unfortunately, he ascribes it to “some philosophers” without further specification:

> Is it a necessary condition for having [propositional] knowledge, not only that what one is said to know should in fact be true, but also that one should be performing some special mental act? Is it perhaps a sufficient condition, or even both necessary and sufficient? Some philosophers have maintained not only that there are such cognitive states, or acts, but that they are infallible. According to them, it is impossible for anyone to be in such a state of mind, unless what it purports to reveal to him is really so. (Ayer, 1956, 9-10)

But there is an indication that he took the view to be more than a recent Oxford fashion. He writes later on:

> The quest for certainty has played a considerable part in the history of philosophy: it has been assumed that without a basis of certainty all our claims to knowledge must be suspect. (Ayer, 1956, 41)

and adds that many philosophers have taken certainty to require “conditions which exclude not merely the fact, but even the possibility, of error” (Ayer, 1956, 44). As a consequence, he argues, they held that only *a priori* or necessary statements could be known. Ayer may have each or all of Aristotle, Plato, Descartes, and Hume in mind here — in addition to his own earlier self (1936/1990, chap. IV). It is thus plausible that he

---

19. For instance 476e-478e. See Fine (1990) for a criticism of the two-objects view.
shared Prichard’s and Woozley’s idea that the traditional conception of knowledge was that of an infallible mental state.

Like Woozley (1949, 181-184), Ayer criticises the mental state view for having sceptical consequences:

The consequence of accepting [that view] would be that no one could ever properly be said to know anything at all. The reason for this is that there cannot be a mental state which, being as it were directed towards a fact, is such that it guarantees that the fact is so. (Ayer, 1956, 21)

Ayer’s own view is that knowledge does not require such infallibility or certainty. It is sufficient that one has a true belief and the “right to be sure”. In the case of contingent propositions, inductive evidence is sufficient for having the “right to be sure” (Ayer, 1956, 30, 74, 83). This has been (rightly) taken by Gettier (1963, 121) to be a justified true belief analysis which his examples showed to be wrong.

Ayer manifestly did not take his own conception to be a version of a “traditional” one. He rather appeared to have defended his view as an alternative to the traditional one.

Chisholm on his view

The first paper in which Chisholm presented his definition of knowledge opens as follows:

“Evident,” ”unreasonable,” ”certain,” ”know,” ”see,” ”hear,” and ”probable” are instances of epistemic terms - terms which are used in appraising the epistemic, or cognitive, worth of statements, hypotheses, and beliefs. I shall suggest definitions of the more important of these terms, making use of one undefined epistemic locution. (Chisholm, 1956, 447)

Similarly, in Perceiving, Chisholm introduces his definition as something he “suggests” (Chisholm, 1957, 1, 16). He does not present it as traditional or as something that has been defended before. To my knowledge, it is
only after the publication of Gettier’s paper that Chisholm referred to the justified true belief analysis as the “traditional” one.

It is of course possible that Chisholm had taken the analysis to be the traditional one all along. After all, Russell had put forward a formally similar definition as “traditional” in *Human Knowledge and its Limits* (Russell, 1948, 140). But given what philosophers like Woozley and Ayer were writing at the time, it may have not at all been clear to him initially whether the analysis he was defending was a version of a traditional one.

### 2.1.5 How the Legend appeared

The Legend settled somewhere between 1960 and 1967. Gilbert Ryle’s “Epistemology” entry in Urmson’s *Concise Encyclopedia* (Urmson, 1960) ascribes an infallibilist view to modern philosophers and makes no mention of justified true belief. Less than a decade later, in his article on “Knowledge” for Edward’s *Encyclopedia* (Edwards, 1967), Anthony Quinton writes that the justified true belief analysis was the traditional one and that it has been refuted by Gettier’s paper. Here is how I think it happened.

In the 1950s, many philosophers took the traditional conception of knowledge to require some form of infallibility and to have the consequence that almost no contingent truth could be known. Malcolm (1952), Ayer (1956) and Chisholm (1957) considered the consequence unacceptable, rejected the infallibility requirement, and put forward the justified true belief analysis — see sec. 2.5.1 below.

Gettier (1963, 121) was perhaps the first one to point out that Plato appeared to have considered the analysis. Chisholm (1966) subsequently presented it as the “traditional” one. Now, following Gettier’s lead,

---

20. Ryle’s entry can also be read in the re-edition of the *Concise Encyclopedia* (Rée and Urmson, 2004). Ryle’s reconstruction of the traditional view is remarkably close to what I call Ur-foundationalism below.

21. Revision note. This is only true of the 2nd edition of *Theory of Knowledge* (Chisholm, 1977). In the 1966 edition, Chisholm only hints that his definition of knowledge as *evident true belief* is a common one (1). If anything, he presents Russell’s (1948, 170–1) tentative definition of knowledge as true belief *with adequate evidence* as the traditional
Chisholm thought that Gettier-style counterexamples required the assumption that a belief could be justified and false, and he also thought that the tradition rejected the assumption. Nevertheless, he wrote that Gettier’s counterexamples refuted the traditional analysis. In that context, the claim should be read as a shorthand for the following: the infallibilist version of the traditional analysis is ruled out by its sceptical consequences, and the fallibilist one by Gettier’s problem. However, the qualification has been quickly lost, and by 1967, Anthony Quinton simply writes that the traditional analysis has been refuted by Gettier.

If the above sketch is true, then the New Story is not so new. Something pretty close to it was on the mind of those who introduced the justified true belief analysis, but was lost on the following generation of epistemologists.

### 2.2 Ur-foundationalism

One may think that philosophers’ views on knowledge varied so much across history that there is no such thing as “the traditional conception of knowledge”. I do not think so. On my view, there is a traditional conception of knowledge, and it goes as follows.

**Ur-foundationalism**  
S knows that $p$ iff S’s belief that $p$ is an instance of basic knowledge or derived knowledge.  
S’s belief that $p$ is an instance of basic knowledge iff S’s belief that $p$ bears a **discernible and infallible mark of truth**.  
S’s belief that $p$ is an instance of derived knowledge iff S’s belief that $p$ is derived from things S knows by an **inference of a truth preserving kind**.

I spell out the core notions of **discernible and infallible marks of truth** and of **truth preserving inferences** below (sec. 2.2.1, 2.2.2). I then illustrate Ur-foundationalism with a “naive” version of it (2.2.3), which allows me to one (see p. 6). But he insists on how it differs from his definition (because it takes basing into account and allows for basic beliefs). That being said, the idea that Gettier cases arise from allowing evident but false beliefs is already stated (p. 22n).
point out that the discernibility requirement is not by itself internalist, and to suggest that something like Ur-foundationalism may be the ordinary conception of knowledge. I also compare discernibility with related notions in the literature: access and luminosity (sec. 2.2.4, 2.2.5). Finally, I discuss the need for an additional basing condition (2.2.6), and mention the possibility of developing Ur-foundationalism in a coherentist manner (2.2.7). 22

2.2.1 Marks of truth

Marks of truth A belief bears a discernible and infallible mark of truth iff it has a property of a relevant discernible and truth guaranteeing kind, where:

a property of a belief is discernible iff a sufficiently attentive subject believes that a belief of hers has that property iff it has, and

a property of a belief guarantees truth iff any belief that has that property is true.

As the definition makes clear, I do not mean anything special by “mark” and “bearing a mark”. A mark is simply a property of the belief; a belief “bears” a mark if it has the corresponding property. For instance, the truth of a belief is (trivially) an infallible mark of truth. But it is not a discernible mark, since even an attentive subject may believe that one of her beliefs is true while it is not. By contrast, having the content that it rains is thought to be a discernible mark: a sufficiently attentive subject will believe that her belief has the content that it rains iff it has.

To be an infallible mark, it is not sufficient that a property be itself truth-guaranteeing. It needs to belong to a relevant kind of properties all

22. Barnes (1990b, 123-126) sees a “foundationalist” schema in “all Dogmatic epistemologies” (ancient and modern) that is close to what I see as the traditional conception of knowledge. The crucial difference is that his clause for basic knowledge does not include a discernibility requirement, as opposed to mine. His clause for basic knowledge is that one has basic knowledge that \( p \) iff “\( p \) belongs to the class \( \beta \) of basic beliefs”, where different epistemological views give different accounts of the class \( \beta \) (126). This leads him, I think, to overlook the difference between Ur-foundationalist and contemporary externalist theories (see below sec. 2.3.2).
of which are truth-guaranteeing. For instance, *having the content that 331 is prime* is a truth-guaranteeing property of a belief: any belief with that property is true, since 331 is prime. Moreover, it may be a discernible property: prima facie, a sufficiently attentive subject should believe that her belief has the content that 331 is prime iff it has. So if having a discernible truth-guaranteeing property was sufficient for basic knowledge, we would have to say that anybody who merely believes that 331 is prime thereby knows that it is. Here we avoid the consequence by saying that a property is an infallible mark only if *it belongs to a relevant kind* of truth-guaranteeing properties. For instance, *having the content that 331 is prime* may be thought to belong to a kind comprising: *having the content that 332 is prime, having the content that 333 is prime*, and so on. It may be that all the properties of this kind are discernible; thus *having the content that 331 is prime* is a discernible mark. But not all of them are truth-guaranteeing. So if that is the only relevant kind for the property of *having the content that 331 is prime*, then that property is not an infallible mark of truth.

Analogous considerations screen out irrelevant discernible properties. For instance, there may be a specific brain pattern $C$ such that one believes that one’s belief *is realised by pattern $C$* if and only if it is. So *being realised by pattern $C$* turns out to be a discernible property. But is not a discernible mark, if its relevant kinds include properties such as *being realised by pattern $C^*$* where the latter is not discernible.

Which kinds of properties are relevant? Philosophers who held Ur-foundationalist views never raised such questions explicitly. They simply focused on some kinds of properties, ignoring bizarre or gerrymandered kinds. I do not know of any good way to restrict the potentially relevant classes of properties beforehand. So I rather leave it to each particular version of Ur-foundationalist to specify the relevant ones.

The definition also leaves open what a sufficiently attentive subject is. Sufficient attention is implicitly relative to a belief and a mark: a mark of a belief is discernible if a subject sufficiently attentive to *that belief and that mark* cannot be mistaken about the presence of the mark. Thus we need not postulate that sufficiently attentive subjects attend
to all of their beliefs at once. Moreover, sufficient attention requires at least the capacity to form beliefs about one’s beliefs. But that does not mean that subjects who lack such capacities cannot have beliefs bearing discernible marks. Roughly, the belief of an unsophisticated subject bears a discernible mark iff a sufficiently attentive subject would (non-trivially) believe that a belief of hers bears the mark iff it does. For the rest, sufficient attention is left to particular conceptions of discernibility to specify. On some conceptions, sufficient attention is just reflection; on others, it may require some training. But there is an upper bound on what it requires. If only an omniscient god counted as attentive enough, the truth of a belief would become discernible (since the god thinks that a given belief is true if and only if it is), and the definition would count any true belief as knowledge. Sufficient attention should rather embody some sort of human cognitive ideal.

The notion of a discernible mark of a belief can be extended to discernible conditions. A condition is something a subject is in at a certain time: being in pain, being in the rain, being in Paris are conditions. A condition is discernible iff a sufficiently attentive subject believes that she is in that condition iff she is. A condition is of a discernible kind iff it belongs to a relevant kind of conditions all of which are discernible. Thus being in pain is thought by many to be a discernible condition: a sufficiently attentive subject, it is thought, believes that she is in pain iff she is. Classical propositions are a limiting case of conditions: the proposition that 331 is prime is the condition of being such that 331 is prime. A subject is “in” such a condition iff any other subject is “in” it at any time. Thus we can also talk of facts of a discernible kind: the fact that 331 is prime is a fact of a discernible kind iff the condition of being such that 331 is prime is of a

23. The “non-trivially” qualification is meant to rule out things like being the belief of an inattentive subject as discernible marks: if no attentive subjects could believe that their beliefs are those of an inattentive subject, then trivially a attentive subject would believe that she has the belief of an inattentive subject if and only if she has — that is, never. One could raise marginal problems for the characterisation: for instance, one may want to count not being the object of a higher-order belief as a discernible mark that beliefs may have. But such quibbles may safely be left aside here.
discernible kind.  

2.2.2 Truth preserving inferences

An inference is of a truth-preserving kind iff no inference of that kind could lead from true beliefs to a false one (or to a non-true one). Nowadays, we would call such inferences *deductive*. But as we will see sec. 2.3.4, some historical views counted induction as a truth preserving type of inference. The scope of truth-preserving inference was thought to be much wider than what we now take it to be.  

Here as well, not any way of classifying inferences is relevant. If we counted *inference from some beliefs to a true belief* or *inferences from some beliefs to the belief that* 2 + 2 = 4 as relevant kinds of inference, the definition would give nonsensical results. But I doubt whether we can formulate any substantial general constraint on how to individuate kinds of inference. For instance, on Descartes’s view, an inference is deductive iff we can clearly and distinctively perceive that the conclusion must be true if the premises are. There is no constraint on the logical form of premises and conclusion. So here is a well, we leave it for various versions of Ur-foundationalism to specify what kinds of inferences are relevant.

2.2.3 Naive Ur-foundationalism

Prior to philosophical reflection, one could naturally think that a wealth of conditions pertaining to one’s perceived environments are discernible. For instance, it seems that a sufficiently attentive subject cannot be mistaken about whether or not they have an elephant in view in

---

24. Conditions in general correspond to *de se* propositions, and can be defined as sets of centred worlds. See Williamson (2000, 94-95) and below sec. 3.1 for the notion of a condition.

*Revision note*: the notion of condition used in 3.1 is more fine-grained than the one is used here, since our conditions are proposition-centred.

25. Cf. Russell (1948, 140): “deduction has turned out to be much less powerful than was formerly supposed; it does not give new knowledge, except as to new forms of words for stating truths in some sense already known.”
daylight. So *having an elephant in view in daylight* would seem to be a discernible condition. Accordingly, *being a belief of somebody who has an elephant in view in daylight* would appear to be a discernible mark of a belief. One’s belief that there is an elephant nearby, while one has an elephant in view in daylight, would thus bear an infallible and discernible mark of truth. So it is prima facie compatible with Ur-foundationalism that we have basic knowledge of many facts about our environment.

Similarly, prior to philosophical reflection, one could naturally take many inductive inferences to be truth-preserving. Thus one could easily take an inference from *I left the keys in the drawer one hour ago* and *I am alone in the house* to the conclusion that *the keys are in the drawer now* to be of a truth-preserving kind. So it is compatible with Ur-foundationalism that one often gains knowledge through inferences that philosophers would classify as inductive.

It is only reflection on dreams, illusions, hallucinations, and a range of thought experiments that leads us to the idea that facts about our perceived environment are not of a discernible kind and that inductive inferences are not truth-preserving. It is dubious that the consequence is consistently taken into account or even acknowledged outside of philosophical theorising.

This leads me to two remarks.

First, *the discernibility requirement is not by itself internalist*. We will see considerations of discernibility push towards various notions of the “inner” such as sense-impressions and internalist justification. It is thus worthy of note that discernibility does not presuppose such notions: the idea that some “external” facts are discernible is perfectly sound. If anything, the metaphysical notion of the inner grows out of the epistemic notion of discernibility, not the reverse.

Second, *the Ur-foundationalist conception may be the ordinary one*. It may very well be the case that we ordinarily operate with an Ur-foundationalist conception of knowledge coupled with an inflated view of the discernible and the truth-preserving. The New Story does not assume this, but it would certainly be consistent with the widespread acceptance of Ur-
foundationalism in pre-Humean philosophy (and beyond).

### 2.2.4 Discernibility and access

Discernibility is related to what is called “access” in epistemology and to Williamson’s notions of “luminosity” and “cognitive homes” (Williamson, 2000, 24, 93-94). Let me explain how and say why I prefer using the notion of discernibility here.

In epistemology, internalism about justification is often defined as the idea that justification supervenes on facts that are “cognitively accessible” to the subject. This is sometimes glossed as facts the subject is “capable of becoming aware of” (BonJour in Dancy and Sosa, 1992, 132-133). But what does one have “cognitive access” to? What is one “capable of becoming aware of”? Many appear to have straightforward intuitions about this and do not elaborate any further. Those of us who do not are left to reconstrue the notion. Clearly, one is not “aware” of facts one does not even believe to obtain, at least if one would not come to believe them upon reflection either. This is why the proper functioning of one’s brain is said to be something we do not normally have “access” to. But it is also clear from the literature that what we are “aware of” is not everything we know. For instance, even if I can plainly see that it is daylight now, I am not “aware of” the fact that it is daylight, in the proprietary epistemological sense. Sometimes it is said that what one has access to is what one can find out “upon reflection”. 26 But here as well, the term is used with a proprietary epistemological sense. It is natural to say, for instance, that I can realise upon reflection whether I have had butter yesterday. Yet whether I have had butter yesterday is not something I can realise “upon reflection” nor something that is “accessible” to me in the proprietary epistemological use of these terms. One gets a better grip on these notions from their intended use in the New Evil Demon problem (Cohen, 1984). If I was right now the victim of a Cartesian Evil

26. The idea was influentially put forward by (Chisholm, 1977, 16–7), who refers back to Lewis (1929, 19). See also Chisholm (1982, 4–5).
Demon, the intuition goes, I would be just as justified in my beliefs as I am now. The idea that justification supervenes on what one has “access” to is meant to deliver that result. This suggests that what one has “access” to is what one would still know if one were in a canonical Cartesian Demon scenario. But of course, the notion is not meant to be defined in such a way. Rather, it is supposed to follow from some natural notion of “access” that what we have access to survives the Cartesian Demon.

I put forward discernibility as an account of the elusive notion of “access”. The accessible facts are just the discernible facts, as defined above. The point of the Cartesian Demon thought experiments is to show that an attentive subject can mistakenly believe that it is daylight while it is not, for instance. It follows that the fact that it is daylight is not discernible. The same goes for the fact that I had butter yesterday. It is because these facts are not discernible that internalists say that we do not have “access” to them.

The notion of discernibility is flexible enough to cover various internalist notions. Conee and Feldman (2004b) distinguish accessibilism, the claim that justification supervenes on facts one has “access” to, from mentalism, the idea that justification supervenes on mental facts. They want to allow facts about the overall coherence of one’s belief system to matter for justification, even though one does not have “access” to them because one cannot consider all of one’s beliefs at once. But Conee and Feldman do not want to count knowing or seeing as mental states; their definition assumes some internalist notion of the mental. Again, the background intuition is that the mental fully survives the Cartesian Demon. But an epistemologically useful notion of “internal mental states” should not count any inner physical state as mental and should preserve attitude contents. I doubt whether a notion that fills the bill will not be ad hoc. At any rate, Conee and Feldman (2004b) do not give any; they simply assume that the mental survives the Cartesian Demon. By contrast, we can simply define mentalism as the idea that justification supervenes on discernible facts, provided that a subject is sufficiently attentive only if she can attend to all her mental states at once. Knowledge states are not
discernible, since even a subject so attentive can mistakenly believe that they know while they do not, so justification does not supervene on them.

From this perspective accessibilism and mentalism differ only in degree, not in kind: mentalism relies on a stronger idealisation of attentive subjects than accessibilism does. This gives the mentalist a wider range of discernible facts. On the opposite side, some have tried to put pressure on internalist notions of justification by arguing that only a narrow range of facts about one’s present experience and currently entertained thoughts are “accessible” (Greco, 2005). Their arguments make perfect sense on the discernibility account of access: they argue that a sufficiently attentive subject could be mistaken about her own past experience, for instance. On this weaker notion of sufficient attention (one that does not require perfect memory of one’s past experiences), one gets a poorer domain of discernible facts. Thus various views of what one has “access” to can be rephrased in terms of discernibility, with appropriate variations of what sufficient attention requires.

I am not claiming that the debates among mentalism, accessibilism and narrow accessibilism are merely terminological. Various notions of sufficient attention yield various notions of discernibility. The debates are about which notion of discernibility (if any) is apt to play a central role in the theory of epistemic justification. They are not spurious debates about which notion of discernibility is the “real” one.

2.2.5 Discernibility and luminosity

Williamson calls a condition “luminous” if one is always in a position to know that one is in it (Williamson, 2000, 95). 27

On the Ur-foundationalist view, under a few natural assumptions, discernible marks are luminous. Suppose one’s belief $b$ bears a discernible

---

27. Williamson uses the phrase “in a position to know that $p$” as a factive one: if one is in a position to know that $p$, then $p$ is true. Thus strictly speaking, if one was always in a position to know that one was in a certain condition, one would always be in that condition. The definition above should rather be read as: “whenever one is in that condition, one is in a position to know that one is”. 

mark $P$ — say, $b$ is a belief of somebody who is in pain, if we think that pain is discernible. This means that $Pb$ is a fact of a discernible kind: one believes that one has a belief with property $P$ iff one has. Now consider the belief that $Pb$ that one would acquire if one was attentive enough: there is a fact of a discernible kind that entails its truth, namely $Pb$ itself. So if one was attentive enough, one would know that one’s belief $b$ bears the discernible mark $P$. Discernible marks are luminous.

The argument requires a few more assumptions if we take issues of basing into consideration (see sec. 2.2.6). For instance, we take $P$ to be property of a belief of being based on the fact that one is in pain. If one was attentive enough, one would believe that $b$ is based on one’s being in pain iff it was so based. Now consider the belief that $Pb$ that one would acquire if one was attentive enough. $Pb$ itself cannot play the role of a discernible mark for that belief, for it does not satisfy the basing constraint. Rather, we have to assume that being based on the fact that $Pb$ is itself discernible. The belief that $Pb$ that one would acquire if one was attentive enough would be based on the fact that $Pb$. If this latter property is discernible, it would have a discernible and infallible mark of truth. So one would be in position to know that one’s belief $b$ bears the discernible mark $P$. Again, discernible marks are luminous. Authors in the tradition appear to commonly assume that basing facts are discernible or luminous, so I think they would endorse the foregoing argument.

By similar reasoning, Ur-foundationalism implies that knowledge is luminous, that is, if one knows, one is in position to know that one knows. Suppose one’s belief $b$ is an instance of knowledge. This implies that $b$ bears a mark $P$ that is discernible and infallible. But now, by the reasoning above, the fact that $b$ bears $P$ is discernible. And the fact that $b$ bears $P$ entails that $b$ is knowledge. From this it follows that if one forms the belief that $b$ is knowledge on the basis of $b$’s bearing the mark $P$ (and if it discernible that one forms that belief on that basis), then one knows that $b$ is knowledge. So knowledge is luminous, if we grant the assumption that basing facts are discernible, and if we assume additionally that a sufficiently attentive subject will form the belief that they know on the
basis of their beliefs bearing marks that are in fact infallible.

Most importantly, the luminosity of knowledge does not require that the infallibility of a mark be itself discernible. It need not be the case that a sufficiently attentive subject would believe that \( P \) is infallible if it was; even less that for each property \( P, P', P'', \ldots \), in a relevant range, a sufficiently attentive subject would believe that it is infallible iff it is. To illustrate, suppose that being clear and distinct is a discernible and infallible mark of beliefs, and that one has a clear and distinct belief that \( p \), so that one knows that \( p \). Suppose further that one believes that one knows that \( p \) on the basis of the fact that one's belief that \( p \) is clear and distinct. One's higher-order belief that one knows that \( p \) bears a discernible and infallible mark of truth: namely, it is based on the fact that one's belief that \( p \) is clear and distinct. It is discernible that it is based on that fact (we assume basing facts are discernible), and that fact entails that one knows that \( p \) (since clarity and distinctness are discernible and together infallible). So one knows that one knows. There is no need for an independent mark of the infallibility of clarity and distinctness. The subject need not even believe that clarity and distinctness are infallible.

As a result, Ur-foundationalism tolerates epistemic circularity. An Ur-foundationalist may say that some property of beliefs — say, clarity and distinctness — is a discernible and infallible mark in virtue of which we gain knowledge, and allow that we come to know that the property is discernible and infallible in virtue of our beliefs bearing that very mark. (Either because our belief that the mark is discernible and infallible bears the mark itself, or because it is adequately derived from beliefs bearing the mark.) One can both know and know that one does in virtue of a mark, without independent or antecedent grounds to believe that the mark is infallible.\(^\text{28}\)

\(^{28}\) On epistemic circularity, see Van Cleve (1979), Stroud (1984, ch.1), Alston (1986), Sosa (1997b), Pryor (2000, 524-530), Cohen (2002), Wright (2004). Van Cleve attributes to Descartes a epistemically circular foundationalism that corresponds closely to the one defended by our imaginary Ur-foundationalist. The idea that Ur-foundationalism tolerates epistemic circularity may explain why dogmatic Ur-foundationalists were not moved by Sceptical arguments relying on Agrippa’s trilemma: see sec. 2.3.2 below.
The luminosity of a condition does not imply the luminosity of its negation. It may be that when one is in that condition, one is always in a position to know that one is, while when one is not in that condition, one is not always in a position to know that one is not. Discernibility has the same asymmetry. For a condition to be discernible, it is required that when one is not in the condition, one does not believe one is; for its negation to be discernible, it is required that when one is in the condition, one believes that one is not. It may be discernible that one’s belief has a given mark without being discernible whether one’s belief has a given mark. Consequently, while Ur-foundationalism naturally implies that one is always in a position to know that a discernible mark is present, it need not imply that one is always in a position to know that the mark is not there. Thus some versions of Ur-foundationalism may hold that if one knows, one is in position to know that one does without holding that if one does not know, one is in position to know that one does not. I do not know whether any historical version of Ur-foundationalism emphasised this feature.

If discernible facts are luminous, they form what Williamson calls a “cognitive home”:

There is a constant temptation in philosophy to postulate a realm of phenomena in which nothing is hidden from us. Descartes thought that one’s own mind is such a realm. Wittgenstein enlarged the realm to everything that is of interest to

29. That is, \( p \) is discernible iff for the attentive subject, \( p \rightarrow Bp \) and \( \neg p \rightarrow \neg Bp \). But for \( \neg p \) to be discernible, it must be that \( \neg p \rightarrow B\neg p \).

30. That is, they can endorse “positive introspection” or axiom 4 (\( Kp \rightarrow KKp \)) of epistemic logic without endorsing “negative introspection” or axiom 5 (\( \neg Kp \rightarrow K\neg Kp \)), where \( K \) is read as “being in position to know”. For any given knowledge-providing mark, the following must hold, on the Ur-foundationalist view: if the mark is not there, one is in position to avoid a mistaken belief that it is there. Suppose one has a false belief that \( p \). That belief does not bear any knowledge-providing mark, since it is false. Thus if one was attentive enough, one would not believe that the belief bears one of these marks. But it does not follow that if one was attentive enough, one would not believe that one knows that \( p \). For it may be that one mistakenly believes that some other mark that the belief has is infallible, for instance. The derivation of the luminosity of knowledge does not assume that the infallibility of marks is discernible, as we pointed out.
philosophy. That they explained this special feature in very different ways hardly needs to be said; what is remarkable is their agreement on our possession of a cognitive home in which everything lies open to our view. Much of our thinking — for example, in the physical sciences — must operate outside this home, in alien circumstances. The claim is that not all our thinking could be like that.

To deny that something is hidden is not to assert that we are infallible about it. Mistakes are always possible. There is no limit to the conclusions into which we can be lured by fallacious reasoning and wishful thinking, charismatic gurus and cheap paperbacks. The point is that, in our cognitive home, such mistakes are always rectifiable. Similarly, we are not omniscient about our cognitive home. We may not know the answer to a question simply because the question has never occurred to us. Even if something is open to view, we may not have glanced in that direction. Again, the point is that such ignorance is always removable. (Williamson, 2000, 93-94)

Like cognitively domestic facts, discernible facts are not such that one is omniscient about them, nor that we cannot be mistaken about them. But they are such that we are always in a position to avoid mistakes about them.

Williamson characterises luminous conditions in terms of what one is in a position to know. I have characterised discernible conditions in terms of what one would believe if one were attentive enough. The notions of luminosity and discernibility are thus distinct. There are advantages and drawbacks in using the latter. The major advantage is that Ur-foundationalism can be formulated as a non-circular analysis of knowledge. If we

31. Barring exceptional cases, what one is in a position to know coincides with what one would know if one was attentive enough. “If one is in a position to know p, and one has done what one is in a position to do to decide whether p is true, then one does know p.” (Williamson, 2000, 95).
relied on Williamson’s notion of luminosity, we could only say that the
Ur-foundationalist conception of basic knowledge is that it is knowledge
based on a mark whose presence one is always in position to know.
The main drawbacks are two. First, we need to assume some notion of
relevant kinds of properties (sec. 2.2.1). Second, the idea that discernible
facts are known may face sophisticated counterexamples. Cases might
be set up in which attentive subjects always get things right about a
domain of facts and yet there is no intuition at all that they know such
facts. (Perhaps a case in which even attentive subjects engage in wishful
thinking but the gods systematically fulfil their wishes.) Williamson’s
notion of luminosity automatically screens away such cases, since there
would not be any intuition that the subjects are in position to know those
facts. Still, in all but a few recherché cases, luminosity and discernibility
coincide. So the coarser notion of discernibility captures well enough
the notion of cognitive home while being more informative about the
condition under which one is supposed to know.

2.2.6 Basing

One may worry that Ur-foundationalism, as defined, is too generous
on basic knowledge. Consider the following case:

S is in pain, and believes that she is. However, she only
believes it because her crystal ball says so. (Her attention is
so focused on the ball that she does not introspectively notice
the pain, as it were.)

Now it seems that there is a fact discernible to S that guarantees the truth
of her belief, namely, that she is in pain. Still we would not want to say
that she knows that she is in pain, for her belief is not adequately based on
the fact. Thus, the thought goes, we should also require that the subject’s
belief be based on the fact that her belief bears the mark.

The problem is real, but it does not require a modification of the
clause for basic knowledge. The basing relation should be included in
the relevant properties of beliefs. For instance, instead of being the belief
of a subject who is in pain, we should rather take the relevant mark to be being a belief based on the fact that one is in pain. This undoubtedly creates issues with discernibility (would an attentive subject not only believe that she is in pain, but that such-and-such a belief is based on that fact?) and packs even more into the restriction to relevant kinds of properties. But again, most historical Ur-foundationalists simply ignore such issues. They assume without discussion that some relevant basing relation is satisfied. We will supplement their accounts with such a requirement wherever needed, without getting any further into the problems it may raise.

2.2.7 Coherentist Ur-foundationalism

Paradoxical as it may seem, Ur-foundationalism can be developed in a coherentist manner. Given any notion of coherence of belief systems, a belief may have the corresponding property of being a member of a coherent belief system. Necessarily, if a belief has that property, all the other beliefs of the system have it. It is a property that beliefs can only satisfy in clusters. If the property is said to be discernible and infallible, we get a coherentist version of Ur-foundationalism.

In later Stoic epistemology, the notion of “absence of impediment” may be seen as a step in that direction (see below 2.3.2). Russell’s sketch of a coherentist epistemology — which he presents as the way Hegel should have developed his own view — can be seen as a version of coherentism Ur-foundationalism:

“The coherence theory in its extreme form maintains that there is only one possible group of mutually coherent beliefs, which constitutes the whole of knowledge and the whole of truth.”

(Russell, 1948, 172)

If the coherence of one’s beliefs is discernible, and if the unique coherent set of beliefs must be true, then coherence is an infallible and discernible mark of truth. (Like Russell, I am unable to tell whether Hegel held such a view.)
Now suppose that the coherence of a belief system ensured that *most* of its members are true, without ensuring that all are. This would lead to a *holistic* notion of knowledge. Roughly: even though we cannot say of any particular belief member of the system that it is knowledge, we could say that the system somehow *contains* knowledge, since it is overall guaranteed to be mostly true. On this sort of view it would make less sense to talk of knowledge as instantiated by particular beliefs, as we commonly do. Knowledge would rather be a property of belief systems or theories. I am not sure whether this position of the logical space has been occupied, but Davidson (1983/2001) comes close.

### 2.3 Historical illustrations

The Stoics and Descartes are the most straightforward historical examples of Ur-foundationalist views. These are not random picks. Descartes’s influence on modern epistemology is well-known and still visible. But the place of the Stoics in Ancient epistemology and beyond is no less prominent. I mainly focus on the latter. Descartes’s ideas are more familiar; hopefully the reader can already see how his views fit into the Ur-foundationalist pattern — as we briefly point out in sec. 2.3.3.

Section 2.3.1 argues that the Stoic view of basic knowledge was Ur-foundationalist. Section 2.3.2 rejects the idea, suggested or argued for by Long and Sedley (1987, 251–252), Frede (1987, 157–161) and Barnes (1990b, 131), that Stoic epistemology was externalist. Section 2.3.3 gives an overview of the Stoic conception of derived knowledge and further historical illustrations of Ur-foundationalism: Aristotle, Epicurus and Descartes. Section 2.3.4 finds a confirmation of the widespread endorsement of Ur-foundationalist view in Newton’s and Mill’s idea that induction was a truth preserving kind of inference.
2.3. Historical illustrations

2.3.1 Stoics on basic knowledge

Note. For Hellenistic citations I use Long and Sedley’s (1987) translations. I mention their numbering (LS) and, when available, that of von Arnim (1905) (SVF).

Background on Stoic epistemology

To put things in a short, if inexact, slogan: if Aristotle founded the theory of science (Barnes, 1993, xiv), Hellenistic philosophers founded the theory of knowledge. Around 300 B.C., there was intense interest in ordinary perceptual knowledge. Challenges from radical sceptics undoubtedly played a role (Long and Sedley, 1987, xviii; cf. “Anaxarchus and Monimus [...] compared existing things to stage-painting and took them to be like experiences that occur in sleep or insanity”, LS 1D). The founders of the two new schools, Epicurus (341-271) and Zeno of Citium (334-262), both defended the idea that the senses provided a “criterion of truth”, a yardstick with which each opinion could be tested for knowledge. Around the time of the death of Epicurus, Arcesilaus took the head of the Academy (c.273 to c.242) and gave it a distinctively sceptical orientation. For two centuries, the sharp and sustained debate between the ‘New Academy’ of Arcesilaus and Carneades (head from mid-II\textsuperscript{nd} century B.C. to 137) and the Stoa of Zeno and Chrysippus (head 232-c.206) dominated Hellenistic epistemology (Long and Sedley, 1987, 249). It stalled as the great schools of Athens disintegrated around 100. We find it recorded in Cicero’s (106-43) Academica (esp. 2) and in Sextus Empiricus’ (II\textsuperscript{nd} A.D.) Against the professors (esp. 7 and 8), as well as in various other sources — though what remains is but a fragment of what there was. By then, the concepts and doctrines of Stoic and Academic epistemology were assimilated in the philosophical common ground and were poised to have a longstanding influence (Frede, 1987, 176).

Stoics distinguished two knowledge-like states: cognition and knowledge.\footnote{32. See e.g. LS 41A, SVF 1.66. The word translated as “cognition” is katalēpsis, a}
tions which are immune to objections and are reserved to the ideal wise
man, of which they doubted whether there was ever one. Though they
refer to it by the ordinary Greek word for knowledge, we would rather
call what they had in mind “science” or “wisdom”. By contrast, the
Stoic cognition corresponds more closely to what we would ordinarily call
“knowledge”. It is their theory of cognition that interests me here.

Cognitive impressions

Stoic epistemology is thoroughly empiricist: all knowledge comes
from impressions, which are quite literally images imprinted into the mind.
Impressions have content. We can assent to them, that is, believe their
content, or not. Among them Stoics argued that there was a special
class of cognitive impressions. I hold that the Stoics endorsed the three
following theses:

Infallibility thesis. All cognitive impressions are true.

neologism that literally means apprehension or grasping. (Cicero translates “perceptio”.)
“Knowledge” is epistèmê.

33. See LS 41A (SVF 1.66), LS 41B (SVF 1.60), LS 41C, LS 41H (SVF 3.112) and (Annas, 1990, 187-188). “Immune to objections” corresponds to the Stoic phrase “firm and unchangeable by reason”, which Long and Sedley (1987, 257) interpret as being “impregnable to any reasoning that might be adduced to persuade a change of mind”.

34. See Barnes (1980, 204): “The verb ‘epistathai,’ and its cognates ‘epistèmê’ and ‘epistèmôn,’ are not philosophical neologisms; they occur frequently in Greek literature from Homer onwards, and they are there correctly translated by ‘know’ and its cognates.” In the Stoic context, Long and Sedley nevertheless translate epistèmê as “scientific knowledge” (Long and Sedley, 1987, 257).

35. Long and Sedley (1987, 257): “It would be possible to translate katalèpsis by ‘knowledge’ in many contexts”. See also Annas (1990, 184-185, 189).


37. See Frede (1987, 152-157). Unlike Locke or Hume, Stoics did not take impressions to be something we have introspective access to. Cf. Frede (1987, 167): “we have to avoid thinking of Stoic impressions as pictures or images of the world which can be looked at introspectively, with the mind’s eye, as it were, to see whether they have this feature that guarantees their truth. [...] For we have to take into account that impressions for the Stoics are mental states that are identified as highly complex physical states, as we can see from the fact that originally they were conceived quite literally as imprints. [...] There is no suggestion that we could observe them to find out exactly what they are like.” Our definition of discernibility allow some features of impressions to be discernible without requiring any such introspection.
Discernibility thesis. If one is attentive enough, one believes that one has a cognitive impression if and only if one has.

Ur-foundationalism. One has basic cognition that \( p \) iff one assents to a cognitive impression that \( p \).

One assents to a cognitive impression that \( p \) iff one has a cognitive impression that \( p \) and one believes that \( p \) on the basis of that impression. If the Discernibility thesis holds, Stoics took having a cognitive impression to be discernible. Assuming basing facts and contents to be discernible, being based on a cognitive impression that \( p \) is a discernible mark of truth of one’s belief that \( p \). By the infallibility thesis, it is also an infallible one. Stoic Ur-foundationalism is the idea that being based on a cognitive impression is a discernible and infallible mark of truth.

Cognitive impressions were standardly defined as follows:

A cognitive impression is one which arises from what is and is stamped and impressed exactly in accordance with what is, of such a kind that could not arise from what is not. (LS 40E, SVF 2.65; see also LS 40C, SVF 2.53; LS 40D)

Since cognitive impressions are “impressed exactly in accordance with what is”, they are true (Frede, 1987, 164). So the infallibility thesis holds. Cognitive impressions are also said to be “of a such a kind that could not arise from what is not”. As other texts make clear, Stoics meant by this that cognitive impressions were discernible from non-cognitive ones:

‘Of such a kind as could not arise from what is not’ was added by the Stoics, since the Academics did not share their view of the impossibility of finding a totally indiscernible [but false] impression. For the Stoics say that one who has the cognitive impression fastens on the objective difference of things in a craftsman-like way, since this kind of impression has a peculiarity which differentiates it from other impressions, just as horned snakes are different from others. (LS 40E, SVF 2.65)

It is not entirely clear which discernibility thesis the Stoics held. The two extreme readings are as follows. (1) There is a unique character C
that all cognitive impressions share, and a sufficiently attentive subject believes that an impression has iff it has. For instance, all and only cognitive impressions are “clear and distinct” (LS 40C). (2) Each cognitive impression has some character \( C \) that no non-cognitive impression has, and a sufficiently attentive subject believes that an impression has \( C \) iff it has. For instance, when one looks at Socrates well enough and in proper conditions, one gets an impression of him with a distinctive character that no impression of something else has, and a sufficiently attentive subject believes that an impression has that character iff it has. Various intermediate views are possible. I shall not try to adjudicate, but I will assume the first option for simplicity and concreteness.38

Assuming that cognitive impressions share a common character, we have some indications of what it was supposed to be. Cognitive impressions were supposed to be “clear and distinct” (LS 40C, SVF 2.53). What Stoics meant by “clear and distinct” is not made explicit, however.

38. Call the two readings unitarist and particularist, respectively. In support of the particularist reading, we can point out the fact that Stoics argued for the discernibility thesis on the basis of their metaphysical view that non-identicals were discernible (“no hair or grain of said is in all respects of the same character as another hair or grain”, LS 40J). They seem to conclude from the discernibility of non-identicals that any adequate (that is, detailed enough) impressions of two distinct objects would themselves be distinct. In support of the unitarist reading, some passages characterise the cognitive impressions in general as being “clear and distinct” (LS 40C, SVF 2.53) or “self-evident and striking” (LS 40K). Both Long and Sedley (1987, 250) and Frede (1987, 162) adopt the unitarist reading, but without explicitly considering alternatives. Frede (1987, 162) writes: “the Stoics also seem to assume that cognitive impressions by themselves differ from all other impressions, that there is some internal characteristic that serves to mark them off from other kinds of impressions and allows the mind to discriminate between cognitive and non-cognitive impressions [...]” He interprets the “horned snake” metaphor above as an analogy for “some internal differentiating mark”, by which he appears to mean a unique mark shared by all cognitive impressions. Long and Sedley (1987, 250) adopt a similar position, and assume that clarity and distinctness is the unique mark, on the basis of LS 40C, SVF 2.53.

39. The fragment suggests that clarity has to do with the impression “arising from what is” and that distinctness has to do with the impression “being exactly in accordance with what is”:

“Of the impressions, one kind is cognitive, the other incognitive. The cognitive, which they [the Stoics] say is the criterion of things, is that which arises from what is and is stamped and impressed exactly in accordance with what is. The incognitive is either that which does not arise from what is, or from that which is but not exactly in accordance with what is: one
2.3. Historical illustrations

were also supposed to be “self-evident and striking” (LS 40K). Saying that an impression is “evident” may mean that its content (or the relevant aspect of its content) is obvious: an impression of an ox is “evident” only if it is evident that it is an impression of an ox (Frede, 1987, 159-160). Alternatively, “self-evidence” and the “striking” aspect refer to the same thing, namely that cognitive impressions force assent and action (LS 40H; LS 40K; LS 40O). If their strikingness is to be distinctive, then it should be somehow special, for Stoics also allowed that some non-cognitive impressions were “convincing” (LS 39G, SVF 2.65).

Even though the details are blurry, the intuitive idea behind the Stoic discernibility thesis is clear enough: when one’s eyes are open, one is in a normal state, close enough and in a well-lit area, one can get an impression of an object of a kind that one could not get otherwise. And they thought that impressions of this kind could not be mistaken. The following passage argues that common sense agrees with Stoics on this point:

“So too, whenever someone is keen to grasp something precisely, he is seen to chase after such an impression of his own accord, as when, in the case of visible things, he gets a dim impression of an object. He strains his sight and goes close to the visible object so as not to go wrong at all; he rubs his eyes which is not clear or distinct.” (LS 40C, SVF 2.53)

One idea would be that clarity is a sort of vividness that results from an impression having been caused normally by an external object (as opposed to dreams, for instance), and that distinctness results from the impression being sufficiently detailed (as when one looks at an object from a close distance). But that is sheer speculation.

By contrast, Frede (1987, 159) glosses “clear” as follows (without textual support either): “under normal conditions we not only have an impression which does not misrepresent things but we have one which represents them clearly, that is, affords us a clear answer as to what kinds of objects we are facing. And under normal conditions we do in fact have a clear view of an object we are confronted with, and we can tell without difficulty what its visual features are. Let us call such an impression “clear” and “evident”.”

Frede’s gloss on “distinct” (which draws on Stoic metaphysics and Stoic-influenced medical doctrines) fits the “sufficiently-detailed” reading: “a cognitive impression of an object will involve a representation of this object which is so articulate that the only object which will fit this representation is the very object the impression has its origin in” (Frede, 1987, 162).
and does just everything until he takes in a clear and striking impression of what he is judging, as though he thought the reliability of the cognition rested on this.” (LS 40K)

Thus Stoics thought that when the conditions are appropriate, one can get an impression of Socrates of a kind that one could not get from a twin. They thought that when one is wide awake, one gets impressions such that one could not get while dreaming. More generally, they were committed to the idea that one could always discern whether one had an impression of the cognitive kind. This does not mean that one can always tell whether an impression is true or false; but one can always tell whether an impression is of the cognitive kind or not. And when it is, it is guaranteed to be true. So, surprising as it may be to us, Stoics were committed to the claim that there are some impressions one gets when one perceives that cannot be like the ones one gets in dreams, and that there are some impressions one gets from seeing a certain man that could not be had by seeing his twin (see LS 40I, LS 40C and Frede, 1987, 162–3). Thus, Stoics endorsed the discernibility thesis.

Not only did the Stoics endorse the discernibility thesis, but it was an explicit premise in their debate with Academic sceptics that if there were no discernible class of true impressions, we could not have knowledge. Here is Cicero’s reconstruction:

Zeno defined [a cognitive impression] as an impression stamped and reproduced from something which is, exactly as it is. Arcesilaus next asked whether this was still valid if a true impression was just like a false one. At this point Zeno was sharp enough to see that if an impression from what is were such that an impression from what is not could be just like it, there was no cognitive impression. Arcesilaus agreed that it was right to add this to the definition, since neither a false impression nor a true one would be cognitive if the latter were just such as even a false one could be. But he applied all his force to this point of the argument, in order to show that no impression
arising from something true is such that an impression arising from something false could not also be just like it. This is the one controversial issue which has lasted to the present. (LS 40D, emphasis mine)

Thus Stoics were committed to Ur-foundationalism. In the debate, Sceptics granted the necessity of cognitive impressions for knowledge, but denied that there were any cognitive impressions. (It is unclear what Sceptics are committed to in such debates; they typically accept some of their opponent’s views only for the sake of argument. They appear committed to the necessity of discernibility and infallibility for knowledge, but they might have denied their sufficiency. Stoics themselves thought they were not sufficient for scientific knowledge.)

2.3.2 Were Stoics externalists?

The Ur-foundationalist construal of Stoic epistemology sheds some light on whether and how the Stoic view of basic knowledge is akin to contemporary externalism.

The causal dimension of cognitive impressions

As Frede (1987, 157–61) points out, Stoics typically characterise cognitive impressions in terms of their causal history. It is in virtue of their causal history that cognitive impressions are guaranteed to be true:

[...] it stands to reason that nature has constructed things in such a way that under normal conditions the impression we receive is true. If under normal conditions something appeared to be red or appears to be a human being, then it is red or is a human being. Thus impressions with the right kind of history cannot fail to be true. (Frede, 1987, 157)

This is reflected in the canonical definition of cognitive impressions as “arising from what is.” (LS 40E, SVF 2.65, quoted above p. 2.3.1; LS 40C, SVF 2.53; LS 40D; see also the references given by Frede, 1987, 158 and
by Barnes, 1990b, 132–6). Seen in this light, the Stoic doctrine of cognitive impressions appears close to a paradigm contemporary externalist view, namely Goldman’s (1967) causal theory of knowledge, according to which, roughly, one knows that $p$ iff one’s belief that $p$ is caused by the fact that $p$ in the appropriate way.

As Frede points out, the causal characterisation of cognitive impressions would imply that cognitive impressions are not discernible from non-cognitive ones:

Now normal impressions in general and perceptual impressions in particular have been characterised in such a way that their normality or perceptuality is a relational feature of these impressions, a feature which these impressions do not have by themselves, but only in virtue of the fact that they stand in a certain relation to the world. Hence it would seem that to determine whether an impression is cognitive or perceptual it will not suffice just to consider the impression by itself; we also have to consider its relation to the world. (Frede, 1987, 162)

But crucially, Stoics thought that the causal history of cognitive impressions gave them a specific internal character that was discernible:

But the Stoics also seem to assume that cognitive impressions by themselves differ from all other impressions, that there is some internal characteristic that serves to mark them off from other kinds of impressions and allows the mind to discriminate between cognitive and noncognitive impressions without having to consider their relation to the world. (Frede, 1987, 162)

So while Stoics required an adequate causal history for basic knowledge, they held that an adequate history resulted in a discernible mark. This sets their view apart from contemporary externalist views (see 2.5).
The “no-impediment” clause

Another apparent externalist trend in Stoic epistemology comes from a later addition of a “no-impediment” clause to the Stoic definition of basic knowledge. The theory of cognitive impressions was put under pressure by Sceptics, who argued that putative cognitive impressions were not discernible from some false ones. Later Stoics replied by amending the clause for basic knowledge. They said that one knew if one had a cognitive impression which was “unimpeded” or “undiverted” (LS 40K). The idea of “undiverted” impressions was in fact taken over from the Academician Carneades (see sec. 2.4.2 below and Allen (1994) on Carneades’s own views).

Some texts suggest that the relevant “impediments” consist in the dysfunction of sense organs or an unfavourable position of the subject:

For a cognitive sense-impression to occur, e.g. one of sight, five factors in their [the Stoics’s] view must concur: the sense-organ, the sense-object, the place, the manner and the mind; since if all of these but one are present (e.g. if the mind is in an abnormal state), the perception, they say, will not be secured. For this reason some said that the cognitive impression is not a criterion universally, but when it has no such impediment. (LS 40L)

By contemporary lights at least, that one’s organs function properly and that one is properly located (e.g., not located in an area where the light is misleading) are not discernible conditions. Rather, they are paradigmatic examples of “external” factors.

These texts lead to the following picture: later Stoics would have granted that some false impressions are indiscernible from cognitive ones. Still, they would insist, cognitive impressions differed by their causal etiology: they “arise from what is” in appropriate conditions. Being infallible, they would be sufficient for basic knowledge, even though one could not discern whether one had a cognitive impression or not. In short, later Stoics would have been externalists.
Long and Sedley (1987, 251-252) ascribe this rejoinder to the Sceptics to later Stoics. On their view, all the Stoics need to hold against Sceptics is that *when everything is normal, sense impressions are truthful*. They conclude that “the Stoics had the better of the argument.” To me, that would rather be a major concession on the part of the Stoics. The discernibility of cognitive impressions from non-cognitive ones was a central tenet of the early doctrine.

However, I do not think that the externalist reading is defensible. Cicero, for instance, writes that “Zeno was sharp enough to see that if an impression from what is were such that an impression from what is not could be just like it, there was no cognitive impression.” (LS 40D, quoted above p. 2.3.1) If later Stoics had given up the discernibility requirement, they would have rejected that claim, and Cicero would certainly have been aware of it. He would not have written that the issue of the discernibility of cognitive impressions was “the one controversial issue which has lasted to the present” (LS 40D). Sceptics would undoubtedly have taken note of the concession, and Sextus would have reported it. In short, if Stoics had given up the discernibility thesis, we would have more direct evidence of it. Moreover, it is entirely unclear how the Stoic could go on holding that the wise man never assents to an incognitive impression (LS 41D, LS 41G), unless they implausibly held that the wise man is always in normal conditions.

The “no-impediment” clause should rather be interpreted as requiring *the absence of discernible defeaters*. Other texts about “impeded” or “diverted” impressions suggest that “impediments” consist in opposite beliefs or impressions:

Thus when Heracles stood before Admetus, having brought Alcestis back from the dead, Admetus then took a cognitive impression of Alcestis, but did not believe it . . . for he reasoned that Alcestis was dead, and that one who is dead does not rise again though certain spirits do sometimes roam around . . . . Therefore the cognitive impression is not the criterion of truth unconditionally, but when it has no impediment.
Admetus has a cognitive impression that \( p \) (e.g., that Admetus stands in front of him), but his other beliefs suggest that \( p \) is false, so he does not know that \( p \).

The following text describes Carneades’s notion of “undiverted” impressions, which the Stoics took over, and gives a more detailed idea of what the notion amounts to:

[...] since an impression never stands in isolation but one depends on another like links in a chain, a second criterion will be added which is simultaneously convincing and undiverted. E.g. someone who takes an impression of a man necessarily also gets an impression of things to do with the man and with extraneous circumstances — things to do with him like his colour, size, shape, motion, conversation, dress, foot-wear; and external circumstances like atmosphere, light, day, sky, earth, friends and everything else. So whenever none of these impressions diverts us by appearing false, but all with one accord appear true, our belief is all the greater. For we believe that this is Socrates from is having all his usual features — colour, size, shape, conversation, cloak, and his being in a place where there is no one indiscernible from him [...]. (LS 69E)

The text suggests that whether an impression is “undiverted” depends only on other impressions the subject has. Together, both texts suggest that an impression is “diverted” iff one has other beliefs or impressions that suggest that that impression is false. Thus defined, “undivertedness” could have been a condition that later Stoics took to be discernible.

The problematic texts that present “impediments” as a matter of cognitive dysfunction or inappropriate environment can be reinterpreted in the light of this. Just as early Stoics took the causal history of a cognitive impression to be reflected in some discernible property of it, later Stoics took cognitive dysfunctions and inappropriate environments to
be reflected in discernible defeaters. Treacherous lighting or deranged organs produce impressions that suggest to one that one’s impressions are misleading. Again, Stoics differ from contemporary externalists by requiring a discernible mark of truth.

Note that Carneades’s theory of “undiverted” impressions is an embryo of coherentism. Suppose that I have an impression of Socrates that I cannot discern from impressions of some other person. If I simultaneously have a discernible impression of a place where that other person cannot be, then the two impressions jointly constitute a discernible and infallible mark of Socrates. Thus, drawing on Carneades’s theory of undiverted impressions, later Stoics appear to have developed a coherentist version of Ur-foundationism along the lines we sketched in abstracto sec. 2.2.7.

Basic knowledge and epistemic circularity

Barnes (1990b, 131) ascribes a form of “externalism” to the Stoics — and Greek Dogmatists in general. On Barnes’s use, an epistemology is “externalist” if it endorses a clause for basic knowledge of the following form, for some \( F \) and any \( S, p \):

If \( S \) believes \( p \) because \( F \), then \( S \) knows that \( p \),

without simultaneously holding that:

\( S \) knows that \( p \) only if \( S \) believes \( p \) because \( S \) believes that \( F \). (Barnes, 1990b, 136–7)\(^{40}\)

Thus an epistemology is externalist if it claims that it is sufficient that some fact \( F \) holds for a subject to know, without requiring that the subject believes that \( F \) holds. Since it is not required that the subject believes that \( F \) holds, it is not required that the subject justifiably believes or knows that \( F \) holds.

Barnes’s “externalism” corresponds to a particular answer to Agrippa’s trilemma. Agrippa’s trilemma is the fact that, when prompted to argue

\(^{40}\) Plausibly, the fact in question is intended to be a type of fact relative at least to \( S \) and \( p \), and thus written \( F(S, p) \). We ignore the detail here.
for any belief, one cannot avoid either relying on unargued premises, arguing circularly, or having an infinite number of premises. 41 “Externalists” hold that:

1. There is basic knowledge — knowledge that does not rely on argument.

2. There are facts in virtue of which one has basic knowledge. We (epistemologists, but also subjects) can tell what they are. For instance, we can say that one knows that there is a tree nearby because one has a clear and distinct impression of a tree.

3. But stating the fact in virtue of which one has basic knowledge is not exhibiting an argument on which that knowledge depends. When saying that one has a clear and distinct impression of a tree, or when saying that one has such an impression and these impressions are infallible, one is not saying that one’s knowing that there is a tree is derived from one’s knowledge of one’s clear and distinct impression or of its infallibility. 42

On this basis, “externalists” endorse the first type of answer to Agrippa’s trilemma. Instead of simply stopping at the assertion of what they (purportedly) basically know, “externalists” can further say how they know it, but the explanation is not a justification from which that knowledge is derived. And typically, the explanation is epistemically circular. 43

Barnes (1990b, 132–6) ascribes the view to Stoics (among others) in virtue of the causal aspect of the notion of cognitive impression. On

41. The trilemma is part of Agrippa’s five ways (Sextus Empiricus, 2000, I.164, I.177). Agrippa was a (non-Academic) Sceptic living somewhere between the Ith century B.C and the IIth A.D. See Barnes (1990a) for an overview the five ways and Barnes (1990b) for a detailed study.

42. Cf. sec. 2.2.5, our claim that Ur-foundationalist positions do not require one to have independent or antecedent knowledge that the marks one’s belief bears are infallible.

43. A similar notion of “externalism” can be found in Stroud (1989). Stroud objects to Van Cleve’s (1979) ascription of such an “externalism” to Descartes. In my view the situation is the same with Descartes as with the Stoics: Descartes is “externalist” in the sense that he accepts basic knowledge and epistemic circularity, but differs from contemporary externalists by requiring the discernibility of the properties on which basic knowledge rests.
the Stoic view, it is sufficient for one to have basic knowledge that one has an impression with the right causal history. It is not required that one believes or knows that one’s impression has the right kind of causal history.

I agree with Barnes that Stoic epistemology has this foundationalist kind of structure. In fact, I think that all Ur-foundationalist positions have this structure. And it is true that the same structure can be found in contemporary externalism. There is a crucial difference, however. Contemporary externalism allows indiscernible knowledge makers; Stoic epistemology does not.

(There is also another point worth stressing. Ur-foundationalist epistemologies in general, and Stoic epistemology in particular, need not deny that knowledge making facts are themselves known. To the contrary, they naturally allow that if one knows, one knows that one does — see sec. 2.2.5. They only deny that, in the case of basic knowledge, knowledge makers are independently or antecedently known. Thus higher-order knowledge is (typically) epistemically circular. This feature is not specific to Ur-foundationalist views, however; it is defended also within contemporary externalism by Sosa (1997b), for instance.)

The significance of the difference can be seen in Barnes’s use of a famous Sceptic simile:

Let us imagine that some people are looking for gold in a dark room full of treasures. It will happen that each will grasp one of the things lying in the room and think he has got hold of

44. The structure is what Sosa (1980) calls “formal foundationalism”. Sosa points out that the structure is compatible both with substantial foundationalism, in the sense of their being a specific “foundational” class of beliefs, and substantial coherentism. We saw that this was true of Ur-foundationalism as well, which can be developed in a foundationalist or coherentist way (2.2.7).

45. See notably Sosa (1997b.a). Sosa insists on the (epistemically circular) higher-order knowledge that an externalist views allows.

46. Barnes (1990a, 223) acknowledges that cognitive impressions are “recognisable” as such. He does not think that the point is particularly important, because Stoics do not require that the subject actually recognises that she has a cognitive impression when she has one. He thus leaves this aspect of the notion aside, and retains only the causal one.
the gold. But none of them will be persuaded that he has hit upon the gold even if he has in fact hit upon it. In the same way, the crowd of philosophers has come into the world, as into a vast house, in search of truth. But it is reasonable that the man who grasps the truth should doubt whether he has been successful. (Sextus Empiricus, 1914, 7.52, trans. Barnes, 1990b, 138.)

The sceptics cleverly compare enquirers into the unclear to archers shooting at a target in the dark. It is reasonable that some will hit the target and others miss it; but it will not be known who has hit and who missed. Similarly, the truth is hidden in a deep darkness. Many statements are shot at it, but which of them agree with it and which disagree cannot be known. (Sextus Empiricus, 1914, 8.325, trans. Barnes, 1990b, 139.)

Barnes uses the simile to illustrate the point that Dogmatists do not know whether their favoured mark is infallible, even if it in fact is. Thus even if they know in virtue of the mark, they are forced to conditionalise their corresponding claims. They can at most say, for instance: “if clear and distinct impressions are infallible, then I know \( p \), but if they are not, I do not.” Thus they cannot make unconditional claims anymore, and in that sense, the Sceptic wins (Barnes, 1990a, 224; 1990b, 139–40; as Barnes points out in the latter, Sextus himself does not use the simile against Stoic “externalism”). Sosa (1997a, 231) reads the simile in the same way, but points out that externalist epistemologies allow knowledge that one’s faculties are reliable, albeit in an epistemically circular way — that is, by using these very faculties. So the Dogmatist wins.

On our view the simile is a straightforward illustration of the idea that belief without a discernible mark of truth is not knowledge. This is the core idea of Ur-foundationalism, and it is endorsed by both Stoics and Sceptics. They only disagree over whether our beliefs bear such marks.

Stoic epistemology shares with contemporary externalism a foundationalist answer to Agrippa’s trilemma. Both views can be supplemented
with the idea that one has epistemically circular knowledge of one’s basic knowledge. But Stoic epistemology crucially differs from contemporary externalism by requiring foundational beliefs to have discernible marks of truths. They would have rejected sceptical arguments based on Agrippa’s trilemma, but they would have endorsed sceptical arguments against contemporary externalism based on indiscernibility considerations. 47

2.3.3 Aristotle, Epicurus, Stoics, Descartes

Aristotle

Aristotle held that all scientific knowledge (epistēmē) is knowledge of principles or knowledge demonstrated from premises one knows (Posterior Analytics, I, 2, 71b20-25). Demonstration is truth preserving. Thus Aristotle endorsed the Ur-foundationalist conception of derived knowledge.

As for knowledge of principles (Post. An. II, 19), the matter is less clear. On the standard “intuitionist” view, induction from observation prepares us for an act of “intuition” (nous) of the first principles. On another (Barnes, 1993, 267-270), induction itself gives us knowledge of the first principles. “Nous” or “comprehension” is not a specific faculty or act, it is simply the name Aristotle gives to the resulting state. Aristotle’s theory of “induction” appears to be an empiricist theory of concept acquisition (Barnes, 1993, 264-265). The acquired concept grounds true judgements on the principles. It is not clear exactly how “induction” goes (see Barnes, 1993, 261-265). Suffice it to say here that there is no indication that induction is fallible nor that it is indiscernible whether one has indeed formed a proper concept. Neither is there an indication that intuition is fallible.

47. There is some evidence that this is just what happened. In Hellenistic texts, Stoics appear much more pressed to reply to indiscernibility-based sceptical arguments than to deal with regress problems. The evidence is weak, however: this may just reflect the fact that Academic Sceptics emphasised the indiscernibility issue while the regress one (which was known at least since Aristotle) was not stressed until the time of Agrippa, by which the Stoic school was less reactive.
2.3. Historical illustrations

fallible or indiscernible.

Epicurus

Epicurus famously endorsed the thesis that “all sense-impressions are true” (LS 16F). He apparently defended it by holding a narrow conception of the content of sense-impressions. Thus vision only directly tells us about colours and “shape at a distance”, for instance (Long and Sedley, 1987, 84). Errors come from judging more than what is strictly given in sense-impressions. What is interesting for us is that Epicurus seemed to have embraced this astonishing thesis as the sole alternative to scepticism:

What is Epicurus’s principle? If any sense-perception is false, it is not possible to perceive anything. (Cicero, *Lucullus*, 32.101, trans. Everson, 1990, 161)

The claim is not surprising if Epicurus held the Ur-foundationalist idea that fallible marks cannot give knowledge (see Everson, 1990, 161-162).

The difficulty is then to recover our substantial knowledge of the world on the basis of narrowly phenomenal basic knowledge (Everson, 1990, 179-180). This was done on the basis of both sense-impressions and “preconceptions”. As with Aristotle, there is no indication that Epicurus allowed the possibility of false or misleading “preconceptions”.  

Stoics on derived knowledge

The Stoic conception of basic knowledge is Ur-foundationalist, as we have seen. Their conception of derived knowledge is more elusive (Frede, 48. Everson (1990, 180) ascribes an essentially fallibilist view to Epicurus on the external world. Epicurus would allow that we know something about the external world (contra Cyrenaics who held that we can only know about our impressions), but allow at the same time that our inferences from sense-impressions to the external world are “vulnerable to error”. But there is no direct textual support for this view; Everson infers it from (1) the fact that Epicurus was not a sceptic about the external world, and (2) the fact that Epicurus does not seem to have a reply to the sceptical objection that our inferences from sense-impressions are not truth-preserving. While the objection is recorded (Striker, 1977, 141), the alleged fallibilist response does not fit well with Epicurus’s strict requirement on a “criterion of truth” to “exclude falsehood” (LS 40B, see Long and Sedley, 1987, 88).
One thing is clear: cognitive impressions are supposed to be the “criterion of truth” (LS 40A, LS 40C, SVF 2.53, a.o.), which means that any truth that is known is known in virtue of them.

Some texts say that cognition — that is, what we would call knowledge — is assent to a cognitive impression (Sextus Empiricus, 1914, 7.151, 8.397). It would follow that all knowledge is basic. This is not entirely implausible, since Stoics held many things that we would take to be abstract to be bodies, and to be perceptible, such as the good or virtue. On this view, Stoic epistemology would be a limiting case of Ur-foundationism with basic knowledge only.

More plausibly, Stoics held in fact that the truth of some opinions could be guaranteed by the truth of perceptual impressions (Frede, 1987, 159, 166), in a way similar to Aristotle’s induction:

If [cognitive impressions] are nevertheless called the criterion of truth, it is because because in an indirect way they also guarantee the truth of all other propositions that are known to be true by human beings. And they do this in the following way. They give rise to general ideas, the so-called common notions which the mind forms naturally on the basis of cognitive impressions and which in turn allows us to have further cognitive impressions. And since cognitive impressions do represent things as they are, the common notions based on them will represent things as they are. (Frede, 1987, 166)

The interpretation makes sense of Chrysippus’s claim that cognitive impressions and common notions are the criterion (LS 48C, SVF 4.273): cognitive impressions are the basic criterion, and common notions are a derived one. Here as well, there is no suggestion that the derivation is fallible.

Descartes.

Descartes evidently endorsed an Ur-foundationalist position. (In fact, a good part of his epistemic terminology, if not notions, seems to come
from Stoic sources, though it may have been in an indirect way. In his *Regulae ad directionem ingenii*, he argues that science comes from two operations of the intellect: intuition and deduction (Rule III). Intuition consists in clearly and distinctively perceiving the truth. Descartes argues that with sufficient training, one can always discern whether one clearly and distinctively perceives the truth, and he assumes that only the truth can be the object of clear and distinctive perception. Deduction is simply the intuition that a certain conclusion must be true if certain premises are. Again, it is not possible to properly deduce a true conclusion from false premises. This basic picture was maintained in his mature works.

### 2.3.4 Induction as truth preserving inference

It may seem surprising to us post-Humeans that induction, in the roughly Baconian sense of inferring general propositions from observations, should be conceived as a truth preserving inference. Yet it was. Here are two striking confirmations.

One is to be found in Newton’s famous general commentary to his *Principia*:

> Hitherto we have explain’d the phænomena of the heavens and of our sea, by the power of Gravity, but have not yet as-

---

49. Parallels between Descartes and Stoics include: clarity and distinctness, *perceptio* (Cicero’s translation of the Greek for “apprehension” or cognition), the avoidance of prevention and supposition, the fact that science (*epistême*), as opposed to knowledge (*apprehension*/*perceptio* in the Stoics, *cognitio* in Descartes), requires resistance to objections or to doubt.

In his *Reply to II^rd Objections*, Descartes writes that “[he has] seen many writings by the Academics and Sceptics” (AT VII 130). If he did not read Cicero’s *Academia* and Sextus’s *Outlines of Pyrrhonism* directly, he at least knew them through various discussions of them, such as Marin Mersenne’s *La Vérité des Sciences* (1625). The Stoic doctrine and vocabulary would have been known to him through these. However, Gaukroger (1992, 590-591) argues that Descartes’s notion of clear and distinct ideas rather draws on Quintilian’s rhetorical theory of “vivid and perspicuous” images. But I find both Gaukroger’s argument against the Stoic lineage (namely, that it was so obvious that the Stoic theory was deficient that Descartes would not have taken it over) and his main argument for the rhetorical lineage (namely, Quintilian’s insistence on the need for vivid images and on the need to convince oneself) weak. Nevertheless, it must be granted that the filiation (if any) of Descartes’s epistemological vocabulary cannot be conclusively established.
sign’d the cause of this power. This is certain, that it must proceed from a cause that penetrates to the very centers of the Sun and Planets, without suffering the least diminution of its force; that operates, not according to the quantity of surfaces of the particles upon which it acts, (as mechanical causes use to do,) but according to the quantity of the solid matter which they contain, and propagates its virtue on all sides, to immense distances, decreasing always in the duplicate proportion of the distances. [...] But hitherto I have not been able to discover the cause of those properties of gravity from phænomena, and I frame no hypotheses. For whatever is not deduc’d from the phænomena, is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferr’d from the phænomena, and afterwards render’d general by induction. Thus it was that the impenetrability, the mobility, and the impulsive force of bodies, and the laws of motion and of gravitation, were discovered. And to us it is enough, that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea. (Newton, 1687, III, Scholium Generale, trans. Motte)

Newton claims to have “deduced from the phenomena” the laws of motion, and that the laws are thereby made “certain”. He does so even while granting that “experimental philosophy” proceeds by inferring particular propositions from the phenomena and “rendering them general by induction”. Thus Newton takes induction to be a truth preserving kind of inference.

A second one is found in J.S. Mill’s System of Logic. The aim of Book III, “On Induction”, is to give conditions for “legitimate” induction (Mill, 1843, III, 1, §1). Mill writes:
In order to get a better understanding of the problem which the logician must solve if he would establish a scientific theory of Induction, let us compare a few cases of incorrect inductions with others which are acknowledged to be legitimate. Some, we know, which were believed for centuries to be correct, were nevertheless incorrect. That all swans are white, cannot have been a good induction, since the conclusion has turned out erroneous. (Mill, 1843, III, 3, §3)

By contraposition, if an induction is a “good” one, its conclusion will not turn out erroneous. It is hard to understand this text otherwise than as endorsing the idea that “good” inductions are truth preserving ones. Mill certainly thought that the validity of induction depended on an objective, substantial feature of our world: its “uniformity”. He did not think like Kant that the uniformity of the world was somehow demonstrable. But he appeared to think that given the uniformity of our world, some inductive methods were infallible.

On the hypothesis that acceptance of Ur-foundationalism was virtually universal until recent times, Newton’s and Mill’s positions are not surprising: if they thought that inductive knowledge was possible, they had to think that induction was truth preserving.

2.4 Hume’s predicament, probabilism and idealism

The idea that we have discernible and infallible marks of truths for our ordinary beliefs was put under pressure by Sceptics since its inception. Three reactions to sceptical arguments have been adopted within Ur-foundationalism. One is the reassertion of dogmatism by denying that the possibilities of error brought forward by the Sceptics were possible. Such views are found in philosophers as diverse as the Stoics, Locke and G.E. Moore. Hume’s defence of scepticism did a lot to block this option, by showing that even if discernible and infallible mark of currently perceived
objects were assumed, most of our purported ordinary knowledge would fall down. Sceptical objections, reassertive Ur-foundationalist dogmatist views and Hume’s extension of scepticism are presented in section 2.4.1.

The second reaction was scepticism, within which originated the extremely influential idea that beliefs could be fallibly justified even while falling short of knowledge. The idea was (perhaps dialectically) put forward by the Academic Sceptic Carneades, and we find it recurring across history in Locke, Peirce, Popper and C. I. Lewis, among others. This is what we may call the probabilist tradition, since Carneades’s term for “convincing” impression (pithanos) was translated as probabilis in Latin by Cicero. We present the tradition in section 2.4.2.

From Locke on appeared a third reaction, idealism. The idea appears in Locke’s reply to Cartesian scepticism, and is manifest in Berkeley’s and Kant’s views: namely, that sceptical arguments will be averted if our ordinary judgements are about ideas and not about things distinct from them. Berkeley and Kant do not give up the Ur-foundationalist standard for knowledge; rather, they think that we can see that it is satisfied once we realise that objects of knowledge are ideas themselves. We briefly discuss idealist views in section 2.4.3.

2.4.1 Sceptical objections to Ur-foundationalism

Ancient Sceptical objections to the discernibility thesis

Academic Sceptics focused their attacks on whether there was a class of impressions that were both discernible from other impressions and infallible. They used what are by now the familiar tools of the sceptical trade, namely a series of cases intended to show that true impressions are indiscernible from false ones:

- fakes, such as a wax pomegranate (LS 40F)
- duplicates, such as two eggs, and twins (LS 40H)
- dreams (LS 40H)
- madness (LS 40H)
Each was put forward as a case in which a putatively cognitive impression could not be discerned from a non-cognitive one. Additionally, Sceptics argued that for any particular property that was supposed to set cognitive impressions apart, a false impression could have the same property. For instance, Carneades pointed out that a false impression could be as “striking” and “self-evident” as any true one (LS 40H).

In more modern terms, the sceptical case against perceptual knowledge may be recast in the form of a dilemma:

**Sceptical dilemma for perception** Either the basis of your putative visual knowledge that $p$ is that you see that $p$, or that you have a visual impression that $p$. The first mark is infallible, but it is not discernible: even if you were attentive enough, you could believe that you see that $p$ while you do not. The second may be discernible, but it is not infallible: even if you would not mistakenly think that you have a visual impression that $p$, it is possible that you have one while $p$ is false. Either way, your visual beliefs lack a discernible and infallible mark of truth.

**Dogmatic Ur-foundationalism: Stoics, Locke, Moore**

In reply, Stoics made a few amendments to their doctrine, but held their ground. They conceded that it was not always possible to distinguish two objects (LS 40I), that it was very difficult to assent to cognitive impressions only, because of our precipitation and lack of wisdom (LS 41D, SVF 2.131), and that some training was required (LS 41G, SVF 3.548: “Just as a mother can distinguish between her twins by the habit of her eyes, so you will do if you practise”). Later Stoics also added Carneades’s “no-impediment” clause to their condition for basic knowledge (LS 40K). But they maintained that there was a class of discernible and infallible impressions.

To many philosophers, the moral of the debate was that it was simply incredible that one could have the rich and detailed impressions we normally have in perception without actually perceiving. That may seem
The Legend of the Justified True Belief Analysis

obviously wrong to us now, who are familiar with Evil Demons and Brains-in-vats and who have a much more generous view of what is possible. But it was not so for Stoics and many later philosophers. Thus a couple of centuries later, the Stoic Epictetus writes:

it is the nature of the mind to assent to truths, to dissent from falsities, to suspend judgement with regard to what is unclear. ‘What’s the evidence for that?’ Feel now, if you can, that it is night. ‘Impossible.’ Reject the feeling that it’s day. ‘Impossible.’ Feel or reject the feeling that the stars are even in number. ‘Impossible.’ (Discourses, I.28 2-3, trans. Barnes, 1993, 135)

Epictetus stresses the fact that cognitive impressions force assent, which may suggest a mere pragmatic answer to scepticism: we are not able to suspend judgement anyway, whether our judgements are fallible or not. But he clearly takes assent to be assent to manifest truths: the fact that “truth” and “falsity” are put on a par with “the unclear” suggests that we should read “truth” as “clear truth”, i.e. manifest truth. The underlying idea, then, it is that one cannot have the impressions one has in broad daylight without it really being day.

A similar reaction is found in much later philosophers. Descartes, of course, not only thought like Stoics that nature could not be so treacherous, but he also thought that he could prove it on theological grounds.

Locke held that we have “sensitive knowledge” of the existence of external things we currently perceive (Locke, 1975, IV, 2, §14, 3, §5, §21). He faced the sceptical objection that one could have the idea of something external without there being something external:

But whether there be anything more than barely that idea in our minds; whether we can thence certainly infer the existence of anything without us, which corresponds to that idea, is that whereof some men think there may be a question made; because men may have such ideas in their minds, when no such thing exists, no such object affects their senses. (Locke,
His first and main answer is that the impressions we get in perception are discernible from false ones:

But yet here I think we are provided with an evidence that puts us past doubting. For I ask any one, Whether he be not invincibly conscious to himself of a different perception, when he looks on the sun by day, and thinks on it by night; when he actually tastes wormwood, or smells a rose, or only thinks on that savour or odour? We as plainly find the difference there is between any idea revived in our minds by our own memory, and actually coming into our minds by our senses, as we do between any two distinct ideas. (Locke, 1975, IV, 2, §14)

(His second answer is an ad hominem move that the Sceptic should give up arguing if he thinks that he is merely dreaming that he argues. His third answer, interestingly, is a form of idealism. We return to it in sec. 2.4.3 below.)

A similar reaction is found in Moore’s “Certainty” (1993). Here as well, Moore faces the sceptical objection to the discernibility thesis:

Now, I cannot see my way to deny that it is logically possible that all the sensory experiences I am having now should be mere dream-images. And if this is logically possible, and if further the sensory experiences I am having now were the only experiences I am having, I do not see how I could possibly know for certain that I am not dreaming. (Moore, 1993, 194)

And his answer is to reassert discernibility, but on the basis of present and past experiences:

But the conjunction of my memories of the immediate past with these sensory experiences may be sufficient to enable me to know that I am not dreaming. [...] It is certainly logically possible that I should have been dreaming now; I might have been dreaming now; and therefore the proposition that I am dreaming now is not self-contradictory. But what I am
in doubt of is whether it is logically possible that I should both be having all the sensory experiences and the memories I have and yet be dreaming. The conjunction of the proposition that I have these sense experiences and memories with the proposition that I am dreaming does seem to me to be very likely self-contradictory. (Moore, 1993, 194)

In the same page Moore concedes that if it is logically possible that one’s past and current experiences are not an infallible mark of his not dreaming, he does not know that he is not. By present lights — and by the lights of many philosophers in Moore’s time, such as Ayer —, the claim that the past and present experiences of an awaken person logically entail that they are not dreaming is blatantly false. That Moore feels compelled to defend the claim as a necessary condition for knowledge is a striking confirmation of his implicit adherence to the Ur-foundationalist conception.

The Humean predicament

Hume was more successful than any of his predecessors in establishing that most of our beliefs lacked infallible and discernible marks of truth.

50. One could argue that our ordinary dreams do not involve genuine beliefs, so that our having a belief logically entails that we do not have an ordinary dream (Sosa, 2007, ch.1). But even if that is a truth about our ordinary dreams, I doubt that it can be argued that it is logically impossible for dreams in general to involve beliefs. (At any rate Moore’s discussion could be carried out in terms of the possibility of hallucinating.)

51. In Moore (1993, 195–6), T. Baldwin reproduces an earlier version of the conclusion of that paper which Moore thought was somehow flawed. In it he says that if there was a sudden discontinuity between his present sensory experiences and recent past ones, it would undoubtedly be possible that he is currently dreaming, and he would not know that he was not. He infers that while current perceptual experience may not be logically incompatible with error, current and past experience may. (He thus seems drawn to the claim that continuity in experience is a discernible and infallible mark of truth.) He then considers a potential rejoinder on behalf of the Sceptic. But surprisingly, the rejoinder is not the one we expect: he does not consider the possibility of a continuous, consistent dream. Rather, he imagines the Sceptic as objecting that Moore might simply be dreaming that he remembers his past experiences. (His reply to this is elliptic, but seems to be that one’s experience somehow may be incompatible with one’s merely dreaming that one’s remembers past experiences.) So Moore, like Locke and Epictetus, simply cannot find credible that one may have rich, consistent and continuous impressions while not actually perceiving.
Particularly, by switching the attention to beliefs about the past or the unobserved, he managed to show that even if knowledge of our ordinary perceptual environment was granted, Ur-foundationalism left standing very few of what we ordinarily take ourselves to know. The long-lasting influence of his sceptical arguments is, in our view, what finally brought down the Ur-foundationalist doctrine — to which he himself subscribed.

I will not examine Hume’s sceptical arguments in detail here. Hume’s specific contribution was to set apart deductive from non-deductive reasoning (Hume, 2007, IV). He pointed out a number of possibilities that we would not normally think of, as in the following passage:

When I see, for instance, a Billiard-ball moving in a straight line towards another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse; may I not conceive, that a hundred different events might as well follow from that cause? May not both these balls remain at absolute rest? May not the first ball return in a straight line, or leap off from the second in any line or direction? All these suppositions are consistent and conceivable. Why then should we give the preference to one, which is no more consistent or conceivable than the rest? All our reasonings à priori will never be able to shew us any foundation for this preference. (Hume, 2007, IV, §10, p.21)

The possibilities show that the inferences we normally draw about unobserved states of affairs do not necessarily preserve truth. One could make such an inference from known premises to a false conclusion.

Hume’s argument may be set up as a dilemma for the claim that derived knowledge rests on discernible and infallible kinds of inference.

**Sceptical dilemma for induction** Either the basis of your putative derived knowledge that \( p \) is that you inferred \( p \) from its cause or its effect, or it is that you inferred \( p \) from what you take to be its cause or effect. In the first case, the inference is infallible, but it is not of a discernible

---

52. See Greco (2000) for a recent study.
kind: one could believe that one was inferring something from its cause or its effect while one was not. In the second, the inference is of a discernible kind, but it is fallible: one could infer something from what one takes to be its cause or effect, and yet the conclusion be false.

2.4.2 Ur-foundationalist scepticism

Given that most of our beliefs fail to have discernible and infallible marks of truth, Ur-foundationalism leads to a widespread scepticism. Ur-foundationalist assumptions have led a number of philosophers to endorse sceptical positions. Not all of them are ones we would spontaneously think of as being Sceptics. A crucial element in the history of sceptical Ur-foundationalism is Carneades’s theory of discernible but fallible impressions.

Carneades’s doctrine of “convincing” impressions

In response to a Stoic attempt to offer a pragmatic justification of cognitive impressions — namely that without them, we would be left without any basis for action — Carneades developed a theory of “convincing” impressions.

Carneades distinguished impressions which “appear true” and impressions which “appear false”. “Appearing true” comes in degrees of intensity, and impressions that appear true with a high force are called “convincing” — pithanon, which Cicero translated in Latin by probabilis. Convincing impressions provided a fallible criterion of truth:

Of the apparently true impressions, one kind is dim, e.g. in the case of those whose apprehension of something is confused and not distinct, owing to the smallness of the thing observed or the length of distance or even the weakness of their vision; the other kind, along with appearing true, is additionally characterised by the intensity of its appearing true. Of these again, the dim and feeble could not be the criterion;
for since it does not clearly indicate either itself or its cause, it is not of a nature to convince us or to pull us to assent. But the impression which appears true and fully manifests itself is the criterion of truth according to Carneades and his followers. As the criterion, it has a considerable breadth, and by admitting of degrees, it includes some impressions which are more convincing and striking in their form than others. Convincingness, for our purposes, has three senses: first, what both is and appears true, secondly, what is actually false but appears true; thirdly, what appears true, which is common to them both. Hence the criterion will be the impression which appears true — also called ‘convincing’ by the Academics — but there are times when it actually turns out false, so that it is necessary actually to use the impression which is common on occasion to truth and falsehood. Yet the rare occurrence of this one, I mean the impression which counterfeits the truth [i.e., the second], is not a reason for distrusting the impression [i.e. the third] which tells the truth for the most part. For both judgements and actions, as it turns out, are regulated by what holds for the most part. (LS 69D)

An impression is convincing if it makes some proposition \( p \) appear true. It is even more convincing if it is “undiverted” and “thoroughly explored” (LS 69E, see p. 57 above). Convincing impressions are said to be reliable, but they are not defined in terms of their reliability. They are, rather, impressions which somehow ‘claim’ their own truth with a certain intensity. Convincing impressions are fallible, but they could play the role Stoics assigned to cognitive ones in guiding thought and action.

It is unclear whether Carneades put forward the doctrine as a positive one, or only as a dialectical move against Stoics. On the later reading, Carneades is simply showing that on the Stoics’s own views, there are guides for life that do not require cognitive impressions (Long and Sedley, 1987, 459–60). (Stoics acknowledged convincing impressions, cf. LS 39G, SVF 2.65.)
Two points interest us here. First, whether or not he positively endorsed it, Carneades took his “probabilism” to be compatible with the sceptical view that we do not know anything:

I am coming round to my father’s view, which he used to say was Carneades’s in fact, that I do think nothing is cognitive; yet I also think that the wise man will assent to what is incognisive, i.e. will opine, but in such a way that he realises he is opining and knows that there is nothing which can be grasped and cognised. (LS 69K)

In the Hellenistic context, opining is incompatible with knowing. Thus Carneades’s “probabilism”, as it came to be called, was coupled with the view that we do not know anything.  

Second, the category of impressions that are convincing and true, which is mentioned in the text (LS 69D), had no special epistemic significance for Carneades. Carneades’s view brought forward a (perhaps new) epistemic category, that of fallibly justified beliefs. Later Stoics incorporated the category in their epistemology: a less-than-perfect man can assent to merely “reasonable” impressions (LS 40F). Thus both Stoics and Sceptics recognised two positive epistemic categories: knowledge, or infallibly justified belief, and probable opinion, fallibly justified beliefs.  

Sceptics simply took the first category to non-instantiated. But the category of beliefs that were fallibly justified and true did not have special significance to them. The justified true belief analysis was unknown to them.

**Probabilism in Locke, C. S. Peirce, K. Popper and C. I. Lewis**

A number of philosophers endorsed broadly sceptical views akin to (and possibly indirectly derived from) Carneades’s theory of convincing impressions. I will mention four: Locke, Peirce, Popper and C. I. Lewis.

---

53. Correspondingly, the use of the phrase “criterion of truth” in Carneades’s presentation of his “probabilism” should be taken with a grain of salt. See LS 69 F and Striker (1990, 156).

54. Though on the Stoic view, fallibly marked beliefs are not justified in the strict sense, since the wise man does not assent to them. Academic Sceptics themselves only allowed a very weak form of assent to probable opinions.
They illustrate the permanence of Ur-foundationalist ideas across a wide variety of philosophical outlooks.

Locke distinguishes the *certain*, which is the province of knowledge, from the *probable*, which is opinion and falls short of knowledge: “the highest probability amounts not to certainty, without which there can be no true knowledge” (Locke, 1975 IV, 3 §14). Certainty comes in three degrees (IV, 2): intuition, demonstration, and the certitude attached to the existence of things we currently perceive (IV, 2, §14). The scope of knowledge is severely limited, namely to our ideas (or perhaps, in a wider sense, to conceptual truths) and to the existence of particular things we currently perceive (IV, 3, §1, §5). We do not know the existence of anything unperceived, except that of God (IV, 3, §21), and almost no necessary connexions between properties (IV, 3, §9–14) — “our ignorance [is] great” (IV, 3, §22). But the province of ignorance allows for “probable” opinion. As with Carneades, probable opinion is belief based on a fallible appearance of truth:

*probability* is nothing but the appearance of such an agreement or disagreement [between two ideas] by the intervention of proofs, whose connexion is not constant and immutable, or at least is not perceived to be so, but is, or appears for the most part to be so, and is enough to induce the mind to judge the proposition to be true or false, rather than the contrary. (IV, 15, §1)

As with Carneades, it is assumed that appearance of truth is reliable, i.e., probable judgements are “for the most part” true.\(^\text{55}\) Despite a radically different background, Locke combines two modern versions of Ancient Ur-foundationalist epistemologies: a dogmatic conception of perceptual knowledge broadly in line with the Stoics’s, and a sceptical probabilism in line with Carneades’s.

---

\(^{55}\) Locke’s text is difficult to parse. I understand it as follows: probability is the appearance of truth resulting from a “connexion” which is for the most part the case, or which appears to be from the most part the case. On this construal, Locke hesitates between actual reliability and appearance of reliability. See Owen (2007, 418–33) for further discussion of Locke’s views on probable judgement.
Peirce stressed that all science is fallible. Peirce introduced the term “fallibilism”, and defined it as the doctrine "that we can never be sure of anything" or "that we cannot attain absolute certainty concerning matters of fact" (Peirce, 1950, 58–9). Peirce differed from traditional Sceptics in that he considered fallibilism to be compatible with the existence and progress of science. But precisely, he distinguished science from knowledge, and he appeared to do so on the grounds that knowledge required infallibility:

there will remain over no relic of the good old tenth-century infallibilism, except that of the infallible scientists, under which head I include [...] all those respectable and cultivated persons who, having acquired their notions of science from reading, and not from research, have the idea that “science” means knowledge, while the truth is, it is a misnomer applied to the pursuit of those who are devoured by a desire to find things out. (Peirce, 1950, 3)

Peirce thus defends an optimist version of Carneades’s “probabilism”. The human pursuit of truth is necessarily fallible and knowledge is not possible. But science is possible, and it consists in the gradual convergence towards truth.

Following Peirce, Popper also presented his view as a species of fallibilism (1963/1972, 228, a.o.). But Popper put it forward as a theory of scientific knowledge. Moreover, he insisted that (what he called) scientific knowledge was in continuity with ordinary knowledge, and was the object of traditional epistemology — the epistemology of Plato, Descartes, Locke, Hume, Mill and Russell (Popper, 1959/2002, xxi–ii). Now, Popper thinks that “knowledge” is falsifiable: some theories that are instances of scientific knowledge may turn out to be false. Thus on his use, “knowledge” is not factive: one can “know” p even though p is false. Whether

56. I will not try to disentangle fallibility and corrigibility in Peirce’s views here. See Tiercelin (1993) and Levi (1980, ch.1) for some discussion — though what Levi calls “infallibilism” is different from what is called infallibilism here.

57. Revision note. Duncan Pritchard pointed out that not all falsifiable hypotheses
this is a legitimate use or extension of the word “know” is a question that can be set aside. Popper’s substantial point is that the relevant epistemic category for both our (good) ordinary and scientific beliefs is one that includes false beliefs. There is no epistemically relevant category of true scientific beliefs, for instance. This sets Popper’s view apart from the later justified true belief analysis of Chisholm’s and Ayer’s. And this puts Popper closer to Peirce and to Carneades. Popper’s notion of “knowledge” is a (significantly refined and updated) version of Carneades’s notion of probable opinion.

C. I. Lewis (1929) endorses the idea that all beliefs about empirical matters of fact are probable only. But he nevertheless calls them “knowledge”, in violation of the dichotomy between certain knowledge and probable opinion endorsed by the Stoics, the Sceptics and Locke:

The only knowledge of a priori is purely analytic; all empirical knowledge is probable only. In affirming such a view, one assumes a heavy burden of proof; the whole history of the theory of knowledge (unless we go back to Plato) seems to enforce the conclusion that such a conception must inevitably lead to scepticism. Lewis (1929, 309)58

For Lewis, Hume’s sceptical arguments establish that empirical knowledge is at most probable (Lewis, 1929, 323). On his view, any statement about empirical matters of fact implicitly asserts something about “all future possible experience”, and there is always a possibility of these assertion to turn out false (Lewis, 1929, 280–1). But what does Lewis mean by probable knowledge? He does not mean by it, as one would have expected, knowledge on a fallible basis. Rather, he means knowledge of probabilities. On Lewis’s view, we do not know that, say, there is a tree in the garden, but we know that there is probably a tree in the garden:

are falsified, hence the condition of falsifiability is compatible with factivity. My (ill-formulated) point here is that some falsifiable hypotheses have turned out to be false, and yet as far as I can see, Popper would still count them as instances of scientific knowledge. And to my knowledge, Popper never insists on truth being a condition on scientific knowledge.

58. I do not know on what grounds C.I. Lewis thinks that Plato shared his view.
if we thus state, for example, Newton’s law of gravitation as an absolute truth, we must not confuse what is stated with the judgement of any informed and intelligent person who makes the statement. The intent of the judgement is not the statement judged probable, but that it is probable. If in such a case we assert briefly that A is B, our judgement is, “It is probable that A is B” [...]. (Lewis, 1929, 324–325)

The motivation for Lewis’s view is clearly that by switching to contents of the form probably p, he hopes to restore the infallibility of empirical knowledge. First, he points out that probably p may be true even if p is false:

Now a common supposition seems to be that our knowledge of the law of gravitation is invalid if there are facts of nature which do not conform to the law. But if this is probable knowledge, it is a very simple and obvious fact that its validity does not require such conformity. The judgement “A is B is probable” does not require for its truth that A is B; it requires only that this should be genuinely probable. (Lewis, 1929, 324–325)

And secondly, he argues that a proper probabilistic inference preserves the truth of its conclusion:

The probable judgement, if valid, is true. There is no difference in the case of probability-inference between validity and truth. What the judgement “A is probably B” asserts is not that A is B or that any other objective state of affairs (except what the premises assert) holds good. It asserts that “A is B” has a certain probability on the basis of certain data. (Lewis, 1929, 331)

Lewis’s epistemology thus fully conforms to the Ur-foundationalist pattern. We have basic knowledge of sense-data, which serve as their own infallible and discernible marks. From this we derive knowledge through truth preserving inferences. But the major concession to scepticism is that we do not reach knowledge of facts we would ordinarily take ourselves
to know. We only learn that certain claims are probably true according to the data we have. Carneades “convincing” impressions are recast into deductive proofs of probabilistic claims.

(Some remarks about Ayer and Russell. Ayer (1936/1990, 19) seems to have endorsed a view in Carneades’s tradition: “Indeed, it will be our contention that no proposition, other than a tautology, can possibly be anything more than a probable hypothesis.” He does not use the term knowledge, so it is hard to say whether he had a view like Peirce’s or like C. I. Lewis’s in mind. In spite of its title, his *Foundations of Empirical Knowledge* (Ayer, 1940) hardly mentions knowledge either. In it he restates the view that any empirical belief has a merely fallible basis (Ayer, 1940, 39, 43). His main claim is that empirical beliefs can be based on “reasonable”, though not “demonstrative”, inferences (see e.g. 230). Since he grants that these claims may turn out false, he appeared to endorse a version of Carneades’s scepticism, like Peirce, rather than C. I. Lewis’s infallibilist probabilism.

Russell’s analysis of knowledge in *Human Knowledge* (1948) is not altogether clear to me. He puts forward the “true belief supported by adequate evidence” account with some reticence (Russell, 1948, 170–1). The whole thrust of the book is that Humean scepticism is averted once we recognise that there are grades of knowledge and that these grades are degrees of probability. My best guess is that he has a view like that of Lewis’s in mind.)

2.4.3 Idealism

A third Ur-foundationalist reaction to sceptical arguments is *idealism* and its more modern cousin, *verificationism*. I am not claiming that the difficulties of dogmatic Ur-foundationalist are the only reason for the development of idealist views from Berkeley on. But they certainly played a role.

Locke’s main answer to the sceptical claim that perceptual impressions cannot be discerned from dream ones was an outright denial, as we have
seen (2.4.1). But right after, he sketches an idealist response, in case his denial is not granted:

But yet if [one who argues that a dream may produce the same idea] be resolved to appear so sceptical as to maintain, that what I call being actually in the fire is nothing but a dream; and that we cannot thereby certainly know, that any such thing as fire actually exists without us: I answer, That we certainly finding that pleasure or pain follows upon the application of certain objects to us, whose existence we perceive, or dream that we perceive, by our senses; this certainty is as great as our happiness or misery, beyond which we have no concernment to know or to be. (Locke, 1975, IV, 2, §14)

The rejoinder appears to be the following: even if what I call “fire” turns out to be an idea in me, it is no less a cause (or temporal predecessor) of pain, and that is all that matters to us. The suggestion is that (1) even if I am fallible about the existence of fire, I am at least infallible about the existence of my sensation of fire, (2) and perhaps this is all the knowledge I need. The further step, taken by Berkeley (1710, §1) and subsequent idealists, was to claim that “fire” in fact refers to these sensations. We thus turn out to have infallible beliefs about fire itself. Dogmatic Ur-foundationalism is restored through the concession that knowledge is about our ideas only. ⁵⁹

The dogmatic Ur-foundationalist motivation behind idealism is evident in Berkeley’s and Kant’s case. Berkeley (1710, §18) grants sceptical arguments concerning the existence of things without the mind: sensations are infallible and discernible marks of themselves only, and since there is no necessary connexion between things outside and sensations inside (as dreams show), any inference from the latter to the former is fallible. But instead of concluding that we do not know, he concludes that the things we know about are ideas.

⁵⁹. Locke himself says that our knowledge is “conversant about our ideas only” (IV, 1, §1), but this claim is notoriously hard to interpret (see Newman, 2007, 315–7 for some discussion and references).
Kant’s “refutation of idealism” (1781/1998, B274–9) is an attempt to prove that there are things in space. Kant’s idea is that necessarily, if one is conscious of one’s own existence, then one is conscious of things in space. If the proof is successful, then one’s consciousness of one’s existence, which is presumed to be discernible, turns out to be an infallible mark of the existence of things in space. To achieve the proof, though, Kant concedes that “things in space” are just sensations with a spatial “form”. 60

I think that the same motivation can be seen in later version of idealism and in some verificationist and instrumentalist views, but I will not discuss the point in detail here.

As an epistemological project to restore dogmatic Ur-foundationalism, idealism is flawed. One may argue that present sensations are discernible, and can thus serve as infallible marks of objects on the idealist view that objects are nothing but the sensations themselves. It is more difficult to argue that past sensations are presently discernible. A sufficiently attentive subject may believe that they have had a sensation while they did not. 61 And future ones are clearly not. Any discernible mark we presently have for future sensations is no less fallible than any putative discernible mark we presently have for future events. Ur-foundationalist idealism makes no progress on the problem of induction.

The point is complicated by the fact that some philosophers — such as Kant — thought they could prove that there has to be some uniformity between the present and the future. Others, like Mill, thought they could assume it. They hoped, I think, that a suitable uniformity principle

60. Kant undoubtedly took his view to allow a distinction between subjective space, in which sensations are received, and objective space, in which the understanding somehow “locates” them. What is important to him — and what he thinks distinguishes him from Berkeley — is that his Refutation proves the existence of things in objective space. But he takes objective space to be somehow a “form” of our sensations as well. Kant’s doctrine is intricate; but what is important for our purposes is simply that Kant assumed that knowledge of the “external” world had to satisfy Ur-foundationalist standards, and that he thought that once we gave up the idea that the “external” world consisted of things-in-themselves we could see that it did.

61. A number of philosophers thought that knowledge of past experiences was not as problematic as knowledge of the external world (past or present). We have seen that Moore (1993, 194) appeared to assume this. So did Russell (1912, V) and Meinong: see Teroni (2006, 70).
together with current sensations might entail the existence of future sensations. However, it is clear by now that any such principle will be too abstract to entail any particular future sensations. (This affects knowledge of the past and the present, since on idealist views, knowledge of the unobserved present and of the past is typically cashed out in terms of future sensations.) This is what motivated C. I. Lewis, for instance, to adopt a probabilistic view of all empirical knowledge, as we have seen sec. 2.4.1.

2.5 The Gettier problem and the internalism/externalism divide

Ur-foundationalism leads to scepticism. Virtually all our beliefs lack infallible and discernible marks of truth. This may have been particularly made clear to philosophers in the analytic tradition by the mid-XXth century. On the one hand Cook Wilson, Pritchard and Price defended the idea that knowledge was an infallible mental state. Their view, like the Stoics’, implied that one could always discern whether one knew. Moore seems to have been pulled towards the view. Many, like Woozley and Ayer, saw the claim as utterly implausible. On the other hand, Peirce, Popper, Russell (in the Problems of Philosophy), and the younger Ayer argued that whatever discernible marks our ordinary beliefs had were fallible. Their view, like Carneades’s, implied that we did not have ordinary knowledge, but only probable opinion.62 We could at most take our beliefs to bear discernible and infallible marks of probabilistic claims, as C. I. Lewis had argued.

This is the context in which Ayer and Chisholm were led to reject Ur-foundationalism. They preserved the idea that knowledge required a discernible mark of truth, but abandoned the requirement that the mark should be infallible. Truth was therefore needed as an independent

---

62. Unless we called “knowledge” some false beliefs, as Popper’s view implied. But Russell (1912, ch.13) served as a reminder that it would be very unnatural to call a false belief “knowledge”.

2.5. The Gettier problem and the internalism/externalism divide

conjunct. Hence they defended the analysis of knowledge as (fallibly) justified true belief. Far from being “traditional”, the view was a break from most of the tradition. However, predecessors might be found: the later Academician Philo of Larissa, the XIVth century scholastic philosopher John Buridan, and Thomas Reid may have held a similar view, as well as a few others — though I have not been able to find any clear statement of the view, and doubt whether any can be. The view appeared in Malcolm (1952), and was more explicitly put forward by Ayer (1956) and Chisholm (1957). Sections 2.5.1–2.5.2 recount this part of the New Story.

The view was short-lived, as a result of Gettier’s enormously influential paper (Gettier, 1963). The paper prompted three main developments. First, externalist views provided an alternative way out of Ur-foundationism. They gave up the idea that knowledge requires a discernible mark, and preserved instead the idea that knowledge requires an infallible mark of truth. Second, Mixed internalist views tried to preserve the fallibilist internalist idea by saying that knowledge requires both an infallible mark and a discernible mark. But I argue that they face a dilemma: either they face Gettier problems, or they reduce to straightforward infallibilist views. Third, the failure of both Ur-foundationism and fallibilist internalism led many epistemologists to give a prominent place to the notion of justification instead of knowledge. This view of the post-Gettier landscape is presented in sections 2.5.3–2.5.5.

2.5.1 Fallibilist internalism: Malcolm, Ayer and Chisholm

The genesis of the Justified True Belief Analysis can be found in the opening pages of Malcolm (1952). The paper attacks Prichard’s infallible mental state account of knowledge. It opens with a quote from the latter that makes clear that the view entails that knowledge itself is discernible:

We must recognise that when we know something we either do, or by reflecting, can know that our condition is one of
knowing that thing, while when we believe something, we either do or can know that our condition is one of believing and not of knowing: so that we cannot mistake belief for knowledge or vice versa. (Prichard, 1950, 88)

Malcolm targets the discernibility claim: “This remark is worthy of investigation. Can I discover in myself whether I know something or merely believe it?” (Malcolm, 1952, 178). He presents five cases of the “ordinary usage” of “know” and “believe”. The later two are the interesting ones:

(4) You say “I know that it [the Cascadilla gorge] won’t be dry” and give a stronger reason, e.g. “I saw a lot of water flowing in the gorge when I passed it this morning”. If we went and found water, there would be no hesitation at all in saying that you knew. [. . . ]

(5) Everything happens as in (4), except that upon going to the gorge we find it to be dry. We should not say that you knew, but that you believed that there would be water. And this is true even though you declared that you knew, and even though your evidence was the same as it was in case (4) in which you did know. (Malcolm, 1952, 178–9)

Malcolm uses the pair (4)–(5) to reject Prichard’s view:

Was there any way that you could have discovered by reflexion, in case (5), that you did not know? It would have been useless to have reconsidered your grounds for saying that there would be water, because in case (4), where you did know, your grounds were identical.

The “grounds” are treated as what is discernible: what one “could have discovered by reflection”. The discernible marks of the belief are the same in the knowledge case (4) as in the error case (5). It follows that there is no discernible and infallible mark that one knows in (4), contra Prichard. But it also follows that there is no discernible and infallible mark that one’s belief is true in (4). So the belief bears at most a fallible mark of truth.
But the pair of cases also suggests that fallible grounds and truth are sufficient for knowledge. The only stated difference between (4) and (5) is that the belief is true in one case, false in the other. This is precisely the conclusion Malcolm draws:

There is only one way that Prichard could defend his position. He would have to say that in case (4) you did not know that there would be water. And it is obvious that he would have said this. But this is false. It is an enormously common usage of language to say, in commenting upon just such an incident as (4), “He knew that the gorge wouldn’t be dry because he had seen water flowing there that morning”. It is a usage that all of us are familiar with. We so employ “know” and “knew” every day of our lives. We do not think of our usage as being loose or incorrect — and it is not. As philosophers we may be surprised to observe that it can be that the knowledge that \( p \) is true should differ from the belief that \( p \) is true only in the respect that in one case \( p \) is true and in the other false. But that is the fact. (Malcolm, 1952, 179–80)

The last lines implicitly state the Justified True Belief Analysis: good but fallible grounds and truth are sufficient for knowledge. (It is worthy of note that the view comes as a “surprise”; if the Legend was true, the view should have been banal.) The motivation for the view is clear: on the one hand, the sceptical consequences of Ur-foundationalism are unacceptable. On the other, knowledge is a matter of one’s discernible grounds of belief. Since the latter are fallible, infallibility is not necessary for knowledge.  

Ayer (1956) and Chisholm (1956, 1957) came to a similar conclusion. Ayer defines knowledge as true belief of which one has the right to be sure, and Chisholm defines it as true and evident belief. Ayer’s prime

---

63. Malcolm stresses the point: “Now it certainly could have turned out that the gorge was quite dry when you went there, even though you saw lots of water flowing through it only a few hours before. This does not show, however, that you did not know that there would be water. What it shows is that although you knew you could have been mistaken. This would seem to be a contradictory result; but it is not.” (Malcolm, 1952, 180)
and stated motivation is the necessity to avoid the sceptical consequences of the infallible mental state view (Ayer, 1956, 23, see also sec. 2.1.4). In Chisholm’s case, the idea that “evidence” should be discernible — and therefore, internal — is manifestly central. (I should record, however, a little fact that does not square well with the New Story: one illustration Ayer gives of having the “right to be sure” is a paradigmatic case of externalist justification. Yet overall, Ayer appears to conceive the “right to be sure” in terms of internalist grounds. 64)

It is also worth noting that Ramsey seems to have considered a fallibilist externalist account early on. I will not discuss Ramsey’s case in detail here. See Dokic and Engel (2001, §2.1).)

2.5.2 Philo of Larissa, John Buridan, and Thomas Reid: predecessors of the Justified True Belief analysis?

The Justified True Belief Analysis, or fallibilist internalism, may claim a few historical predecessors. One is Philo of Larissa, head of the Academy in the 1st century B.C. Philo defended a positive reading of Carneades’s doctrine of convincing impressions: the wise man could follow them,

64. The problematic passage is the following: “Suppose that someone were consistently successful in predicting events of a certain kind, events, let us say, which are not ordinarily thought to be predictable, like the result of a lottery. If his run of successes were sufficiently impressive, we might very well come to say that he know which number would win, even though he did not reach this conclusion by any rational method, or indeed by any method at all. […] so far as the man himself is concerned, there need not be any difference [between knowing and guessing]. His procedure and his state of mind, when he is said to know what will happen, may be exactly the same as when it is said that he is only guessing” (Ayer, 1956, 31–2). The case is reminiscent of Unger’s gypsy (Unger, 1968, 163), which is meant to be a case of knowledge without (Chisholmian) justification, and of “chicken sexer” cases (Foley, 1987, 168–9). However, Ayer is ambivalent about it: “Normally we do not say that people know things unless they have followed one of the accredited routes to knowledge. If someone reaches a true conclusion without appearing to have any adequate basis for it, we are likely to say that he does not really know it” (Ayer, 1956, 32). Other texts invite an internalist reading: “one is conceded the right to be sure when one has taken every reasonable step toward making sure”(Ayer, 1956, 44). Ayer seems to me pulled towards externalism under the influence of ordinary language philosophy, which leads him to focus on the conditions under which one ascribes knowledge to a third person. But that is a topic for further study.
as long as he did not assert them and was aware that he could be mistaken (LS 68S, cf. Long and Sedley, 1987, 448–9). It is unclear to me whether Philo only defended the idea that beliefs on the basis of convincing impressions were allowed, even though they could be false (as in the ‘optimistic’ probabilist tradition), or whether he actually thought that convincing impressions and truth are sufficient for knowledge. Some think he did (Striker, 1990, 157).

Another is Jean Buridan, the XIV\textsuperscript{th} century scholastic philosopher. As Pasnau (2010, 26–8) points out, medieval commentators of Aristotle’s \textit{Posterior Analytics} were aware that Aristotle’s doctrine of \textit{epistēmē}, translated as \textit{scientia}, implied that most of what we ordinarily take ourselves to know was not \textit{scientia}. Buridan, following others, tried to devise lesser degrees of knowledge or of “grasping the truth” that would be appropriate to natural sciences and to the practical domain — of which a paradigm case was that of a jury. Buridan argues that there is a fallible kind of evidentness (\textit{evidentia}) for beliefs in the natural and practical domains. On the face of it, this looks like a version of optimistic probabilism. But Buridan puts it under a heading about “how we grasp the truth”. The phrase is reminiscent of the Stoics’s “cognition”; and more importantly, it suggests that perhaps a \textit{fallibly evident but true} belief is a way of grasping the truth — a lower one, but still one.

A third potential predecessor is Thomas Reid. Reid is commonly taken to defend a fallibilist conception of knowledge (see e.g. Sosa, 2009, 61, Greco, 1995, 294), for Reid allows immediate knowledge of external

65. The problem is to understand this elliptic report: “Philo says that, as far as the Stoic criterion is concerned (i.e. the cognitive impression), things are incognitive, but so far as the nature of things themselves is concerned, cognitive.” (LS 68T) It is unclear what to make of the claim; in particular, whether he held Arcesilaus and Carneades’s view that nothing was cognitive, i.e. nothing is known.
66. For instance, it would be possible for a judge to act well and meritoriously by hanging a righteous man because through testimony and other documents it sufficiently appeared to him in accord with his duty that that good man was a bad murderer. (John Buridan (1964, II.1, f. 9ra), quoted and translated in Pasnau, 2010, 32.)
67. As Pasnau (2010, 33) notes, Buridan’s conception of “moral certainty” — fallible certainty appropriate to the practical domain — was influential on the development of the legal notion of a conviction “beyond all reasonable doubts”.
objects through perception, even while recognising that perception is fallible. A detailed study of Reid is more than what I can offer here; but I would like to point out a potential misunderstanding about Reid’s fallibilism. Stoics, the paradigm infallibilists, were “fallibilists” as well, in a sense: none of them claimed that perception was infallible. They fully granted (as opposed to Epicurus) that senses often gave us false impressions. Yet they thought that we had perceptual knowledge of external objects. Is that sufficient to say that there were “fallibilists”? No, because they thought that when one had perceptual knowledge of an external object, one had an impression of a discernible kind that one could not have in a situation in which one is mistaken. In a nutshell, Stoics combined source fallibilism with belief infallibilism: in appropriate conditions, a fallible source provides infallible beliefs. It is thus very important, when discussing Reid, to distinguish three views, each of which subscribes to source fallibilism. The first is the Stoics’s infallibilist internalism: in appropriate circumstances, the source provides beliefs with an infallible and discernible mark of truth. The second is infallibilist externalism: in appropriate circumstances, the source provides belief with an infallible mark of truth, though the mark is not discernible. The third is fallibilism in our sense: the source provides beliefs with a fallible discernible mark of truth, and when such a belief is true, one knows. Only in the third case would Reid be a precursor of the Justified True Belief analysis; in the first, he would have remained within Ur-foundationality, and in the second, he would have been a precursor of infallibilist externalism. I leave the question for further study.

2.5.3 The Gettier problem and externalism

Internalist Fallibilism faces the Gettier problem (Gettier, 1963): if there can be a justified false belief, one can deduce a true belief from a justified false one. The resulting belief is true and justified but not knowledge.

Early on, it was noted that Gettier problems appeared to be a consequence of fallibilism. Gettier (1963, 121) pointed out that his cases
assumed that a belief could be justified and false. Chisholm (1977) suggested the same. The point has been emphasised, and argued for, in the recent literature (Sosa, 1985, 239–40, Sturgeon, 1993, 160–1, Zagzebski, 1994, 69, Merricks, 1995, 845). We will present an argument to that effect in the context of modal accounts of knowledge (4.2).

Externalist theories of knowledge appeared in the wake of the Gettier problem (Goldman, 1967; Unger, 1968; Armstrong, 1968; Dretske, 1969). They provide an alternative way out of Ur-foundationalism. They preserve the idea that knowledge requires an infallible mark of truth, but abandon the idea that the mark has to be discernible. Unger (1968, 158) states the view in the most general terms: one knows if and only one’s belief is non-accidentally true. Given Hume’s predicament, whether one’s belief is non-accidentally true is not a discernible property of a belief. But it is an infallible one: trivially, one’s belief cannot be non-accidentally true and false. Such views are in position to avoid Gettier problems.

The fact that on the externalist view, the knowledge making property is not discernible has been the source of recurring dissatisfaction (see e.g. BonJour, 1980, Cohen, 1984, Pollock, 1986, 116–7). Yet everybody has granted that the justified true belief analysis of knowledge was wrong. This has prompted a number of authors to defend what I call mixed internalist views.

2.5.4 A dilemma for mixed internalist views

A mixed internalist view is a view according to which knowledge requires both a discernible mark of truth and an infallible one. By Hume’s predicament, no property can play both roles. So there will be two conditions on knowledge: that the belief bears a discernible (fallible) mark of truth, and that it also bears an (indiscernible) infallible mark of truth. Such views face a dilemma.

---

68. Cf. Chisholm (1977, 103): “if it is possible for some propositions to be both evident and false, then, as we shall see, it is also possible for some a person S to accept a true and evident proposition without thereby knowing that the proposition is true”. See also Chisholm (1966, 23).
One the one hand, the two conditions may be independent in a way that allows accidental coincidence. The problem is illustrated by the *no false assumption* account (Clark, 1963; Harman, 1973; Lycan, 2006):

**No false assumption** S knows that $p$ iff S’s belief that $p$ is justified, and no assumption in S’s justification for $p$ is false.

(What counts as an “assumption” in one’s justification is a delicate issue. But there is a reading of it on which the point we are about to make is true. This is all that is needed here to illustrate the problem faced by mixed internalist views with independent conditions.) Whether S is justified, on the view under consideration, is an internal matter: it depends, say, on whether the subject has a network of experiences and beliefs which bears some adequate evidential relations to the subject’s belief that $p$. It is assumed to be discernible. By contrast, the truth of all assumptions involved in one’s justification is not: an ideally attentive subject could mistakenly believe the assumptions to be true. We may also assume that the truth of all assumptions entails the truth of the belief. The two conditions are satisfied by coincidence in cases such as Ginet-Goldman’s fake barns story (Goldman, 1976, 772–3). The subject is justified in believing that something is a barn by his experience, and he assumes that the visual appearance of the building is not misleading. The assumptions, and others we may naturally take the subject to make, are true. Yet it is a coincidence that the two conditions are satisfied: if the subject had looked at the next building, he would have had a similar justification, but the corresponding assumptions would be false. The subject does not know, and the analysis turns out to be insufficient. The moral of the example is that a view according to which knowledge requires to satisfy two conditions, an infallible and a discernible one, is liable to face Gettier-style counterexamples if the two conditions can be satisfied by

---

69. If one believes $p$ because one believes $j$, for instance, then one presumably assumes the material conditional $j \rightarrow p$. But the truth of $j$ and $j \rightarrow p$ implies $p$.

70. See p. 159 below for a description of the case.

71. Here we assume that the subject’s justification does not involve the assumption that other buildings in the area do not have deceptive appearances. That this holds about some natural notion of assumption is granted by Lycan (2006).
2.5. The Gettier problem and the internalism/externalism divide

coincidence. (I will defend this point further in sec. 4.4.4.)

On the other hand, an analysis on which the presence of an infallible mark entails the presence of a discernible one is just a species of straightforward infallibilism. Suppose, for instance, that having a visual experience that \( p \) that is appropriately caused by the fact that \( p \) is a relevant infallible mark for beliefs that \( p \). Any belief that instantiates the mark instantiates a putative discernible mark, that of having a visual experience that \( p \). Now it may look as if the suggested account of visual knowledge that \( p \) is a mixed internalist one: one should have a visual experience that \( p \) (discernible condition), and the experience should be caused by the fact that \( p \) in the proper way (infallible condition). If the analysis is presented with belief and truth as independent conditions, it may even look as if the fourth condition is a minor clause to deal with bizarre counterexamples — an “anti-Gettier codicil” (Pritchard, 2010). But in fact, the fourth, infallibilist, condition entails the truth and the justification ones. The infallibilist clause is doing the real work; justification is epiphenomenal.

Thus mixed internalist views face a dilemma. Either they put forward a genuine conjunction of conditions, and they are likely to face Gettier-style counterexamples. Or they put forward what amounts to a single infallibilist condition that entails some discernible property, and their account is only superficially different from straightforward externalist views. In the latter case, the debate between externalist and internalist views turns on whether the indiscernible infallible condition on knowledge should also entail some relevant discernible property. (I will briefly argue that it should not in sec. 4.4.1, though a full treatment of this debate is beyond the scope of this work.) I have no proof that a mixed internalist account cannot be both genuinely conjunctive and avoid Gettier problems. But I think that, as a matter of fact, most internalist-inspired answers to the Gettier problem — notably, so-called “undefeasibility” accounts — fall into either horn of the dilemma.

72. See Bergmann (2006) for a recent and extended defence of the externalist view — though Bergmann frames the debate in terms of justification, not of knowledge.
2.5.5 The contemporary focus on justification

The fall of Ur-foundationalism also explains the contemporary focus of much epistemology on justification instead of knowledge. In our view, the historical origin of the idea of justification is the notion of a discernible mark on the basis of which it is legitimate to assent to a proposition. On the Stoic view, the discernible marks that entitled one to assent were infallible, so there was no difference between justified belief and knowledge. In the "probabilist" tradition initiated by Carneades, (qualified) assent was allowed on the basis of discernible but fallible marks. But these were not considered to be enough for knowledge. (Philosophers such as Locke adopted the first view for some domains and the second one for others.) Hume’s predicament shows that only the second way is open to the Ur-foundationalist.

The Justified True Belief analysis was an attempt to reconcile the idea that knowledge is a matter of discernible marks with the plausible view that we ordinarily know many things. Its failure, and the contemporary reluctance to give up common sense ground, made clear that knowledge and discernibility came apart. For those who were attached to the discernibility element in Ur-foundationalism, knowledge started to lose its interest. Many focused on the putative discernible aspects of knowledge — call them knowledge-level justification. These properties were naturally grouped with the similar putatively discernible properties that are found in non-knowledge cases only, such as having statistical or weak evidence in favour of a proposition — call them lower-level justification.

The focus of justification — in its discernible sense — brings together philosophers of two stripes. One are the heirs of the “probabilist” tradition of Carneades, Peirce and Popper. They think that we have no knowledge. But rather than describing themselves as sceptics, they may argue for instance that the ordinary notion of knowledge is not relevant to...
epistemology, precisely because it is unrealistic.\footnote{I suspect that Bayesians such as \textit{Jeffrey} (1965) hold such a view.} The others are the heirs of the Chisholmian fallibilist account of knowledge. They think that some suitable anti-Gettier clause will tell how we ordinary have knowledge. But they think that for most philosophical purposes — \textit{e.g.} accounting for belief, action and their rationality — only discernible aspects of knowledge are relevant.\footnote{This trend is exemplified by evidentialists such as \textit{Conce and Feldman} (2004a).}

Now since \textit{Goldman} (1979), externalists have also laid claim on the notion of justification. From the present perspective, externalism about justification aims at showing that the putatively discernible properties that internalists rely on are not discernible after all. (In my view, they have succeeded, but I do not intend to defend the point here.) This should not obscure the fact, however, that the notion is originally a product of internalist assumptions.

## 2.6 Conclusion

Let me first take stock, and then draw a few lessons of the picture of the history of epistemology I have sketched.

According to a widespread Legend, the Justified True Belief analysis defended by Ayer and Chisholm and refuted by Gettier was the traditional conception of knowledge. The Legend has never been properly defended, and we have pointed out a number of facts that do not fit it. I have put forward an alternative picture of the history of epistemology which also sheds some light on contemporary debates.

On the New Story, there is a traditional view of knowledge, \textit{Ur-foundationism}. \textit{Ur-foundationism} is the idea that knowledge requires a discernible and infallible mark of the truth of one’s belief. We have seen its influence throughout history, from Aristotle to Prichard and Moore. But ultimately, the view leads to internalist scepticism, as Hume showed. Dilatory moves have been attempted by endorsing revisionary views of the contents of ordinary knowledge, such as idealism or C. I. Lewis’s
The Legend of the Justified True Belief Analysis

infallibilist probabilism. But by the mid-XX\textsuperscript{th} century, the sceptical consequences of Ur-foundationalism were widely recognised within analytic philosophy. In the context of Moore’s defence of common sense and of ordinary language philosophy, epistemologists were much more reluctant to endorse them than had been the case historically.

Among those who had endorsed the sceptical consequences of Ur-foundationalism, an influential tradition was the probabilism initiated by Carneades, of which the fallibilism of Peirce and Popper are contemporary instances. These views argue that some discernible but fallible marks of truth warrant some form of assent, even though they do not provide knowledge.

The Justified True Belief analysis was the first attempt to provide an alternative to Ur-foundationalism. Drawing on the probabilist tradition, it preserved the idea that knowledge required discernible marks of truth, but claimed that fallible ones were sufficient, in conjunction with truth. The view was quickly shown wrong by Gettier. The New Story thus explains why Gettier cases were discovered so late. Mixed internalist views try to preserve the outlook of the justified true belief analysis, while avoiding both scepticism and the Gettier problem. They hold that knowledge requires, in addition to a discernible fallibilist mark, an indiscernible infallibilist one. But they assume that the latter plays a secondary role, and that for most purposes, the discernible aspects of knowledge are the relevant ones. I have argued that such views face a dilemma: either they face more complex Gettier problems, or the infallibilist clause renders the justification one epiphenomenal.

Externalist views provide the other alternative to Ur-foundationalism. They give up discernibility instead of infallibility. This puts them in a good position to avoid both scepticism and the Gettier problem. Yet giving up the idea that knowledge has a discernible core has appeared too high a price to pay for many. The New Story thus explains the division of contemporary epistemology into externalist and internalist views. It also explains the current focus of much epistemology on justification, that is, on the putative discernible aspects of knowledge and probable
opinion.

If true, the New Story teaches a number of lessons. The first is the instability of fallibilist internalism about knowledge. Fallibilist internalist views are pulled in opposite directions, and it is doubtful that they can dodge the choice. Either they join the probabilist tradition and acknowledge the split with common sense — that is, scepticism, perhaps made more acceptable by arguing that the unattainable standard of knowledge is not the epistemically interesting one. Or they join the externalist camp. Many epistemologists, I think, are currently avoiding facing the choice by discussing epistemological questions mostly in terms of justification without adopting any substantial stance on knowledge.

The second is the instability of fallibilist externalism. A fallibilist externalist theory of knowledge combines the inability of fallibilism to solve the Gettier problem with the counter-intuitive implications of indiscernibility. The point of externalism, the Story shows, is precisely to preserve the infallibilist aspect of the notion of knowledge. Moreover, by laying claim on justification, externalist theories are on a dialectically fragile ground. The notion of justification is originally motivated by considerations of discernibility. Once it is argued that the putative discernible properties that interest justification-based epistemologies are not discernible, it is unclear why these properties should still have a role to play in epistemology. Externalist theories of justification are in danger of lacking a motivation.

The third lesson, and the only one we will build on here, is that knowledge requires infallibility. The only account of knowledge that did not require some form of infallibility was the Justified True Belief analysis, and it is the only one that is clearly and (almost) uncontroversially shown inadequate. Each of Ur-foundationalism and externalism has unpalatable consequences: scepticism on the one hand, the indiscernibility of knowledge on the other. But none is as clearly off the mark as to what knowledge is as the Justified True Belief analysis is. This provides both a historical background and a motivation for the brand of infallibilism to be defended here.
Part II

Epistemology
This chapter lays the groundwork for the main argument of the thesis. The argument, to be given in the next chapter, is that a successful modal account of knowledge must meet two structural constraints, and that method infallibilism is the simplest account to satisfy them. The argument is partly conditional on the assumption that an account of knowledge in modal terms is desirable. In the present chapter I present some methodological considerations in favour of such an account. The considerations are resistible and I do not attempt to establish that we need an account of knowledge nor that the account has to be a modal one. My attitude is rather to assume the modal approach and to see where it leads. The result should still interest those who are or have been treading this path, as well as those who are confronting similar problems from other directions.

The main argument of the thesis relies on structural considerations. The task of the present chapter is pave the way for these by providing a more rigorous notion of modal requirements on knowledge. First (3.1), we introduce a terminology of cases, conditions and requirements, which extends the apparatus of Williamson (2000, 52). Cases are worlds centred on a subject, a time and, additionally, a proposition. Conditions are coarsely individuated by the cases in which they obtain. Requirements are linguistic or conceptual statements of conditions. They are used to state
necessary and/or sufficient conditions on knowledge.

The distinction between conditions and requirements helps to clarify an important point (3.2.1): a requirement may be infallibilist, that is, it may entail some form of impossibility of error, without being itself formulated in modal terms. Thus even if it is granted that an infallibilist condition on knowledge is required, it does not follows that one needs a modal infallibilist requirement. A requirement is modal, in our sense, if it stated only in terms of belief, truth, some notion of basis of belief and a relevant notion of possibility. Undefeated justification accounts, Plantinga’s proper function account, Greco’s and Sosa’s credit or aptness account give examples of requirements that rely on some other notion than possibility, in addition to belief, truth and some notion of basis of belief. They can be infallibilist without being modal. Nevertheless, there are methodological considerations in favour of modal requirements (3.2.2).

The rest of the chapter (3.3–3.8) lays out the general structure of modal requirements. Modal requirements on knowledge are requirements that state that a certain type of error is in some sense impossible in a case. Such requirements have a common structure. I give three versions of the structure: simple (3.4), parametric (3.5) and double (3.7). Simple requirements are a subclass of parametric ones, which are a subclass of double ones. I also generalise the schema to fallibilist modal requirements, which I call reliability requirements (3.6). Section 3.3 provides an overview of the various schemas.

Most major modal accounts of knowledge of the last four decades fit into the schema. Each result from a setting of four parameters: a grounding condition the target case must satisfy, a type of error to avoid, a respect of similarity that fixes a range of other cases in which error should be avoided, and a proportion of cases in the range in which error is to be avoided (all in the infallibilist case, most in the fallibilist case). Going through each account is admittedly tedious, especially given that I will hardly discuss their motivations and problems. Yet there are significant gains in doing so. First, it shows that there is much less diversity in the literature than
meets the eye. Second, it brings out certain overlooked dimensions of the accounts. One is subject-centring, which we discuss in appendix E: are relevant errors only errors the same subject could have made? Third, it allows us to relate various problems with structural features of the accounts. For instance, accounts according to which the range of relevant error varies according to the proposition believed (variable accounts) are expected to violate deductive closure (3.4.4, see also chapter D). These structural features are the ones we use in the argument of chapter 4. The results of this review of major modal accounts are summarised in section 3.8.

Some notation is introduced as we go along; it is summed up in the List of Symbols (E.3, p. 459). Appendix A gives an overview of the different semantics of subjunctive conditionals that may be used to cash out Nozick’s sensitivity and Sosa’s safety conditions.

3.1 Conditions and requirements

3.1.1 Cases, conditions, and requirements

It will be useful to frame the discussion in terms of proposition-centred cases. (This is a slight modification of Williamson’s (2000, 52) terminology of cases and conditions.)

A case $\alpha$ is possible world with a distinguished subject, time and proposition. The subject is referred to as “one” or “the subject” or “$S_\alpha$”. The time is referred to as “the present”, “$t_\alpha$” or by using the present tense. The world is referred to as “the world” or $w_\alpha$. The proposition is referred to as “the target proposition” or “$p_\alpha$”. It is not assumed that the subject believes the proposition. It is not assumed that the subject exists, for that matter. $^1$ The set of cases includes any arbitrary quadruple of a possible world, a subject, a time and a proposition.

Classical propositions are true at a set of possible worlds. If $w_\alpha = w_\beta$
and \( p \) is a classical proposition, \( p \) obtains in \( \alpha \) iff it obtains in \( \beta \). We may also include de se propositions that are true with respect to a subject and a time. If \( w_a = w_\beta \), \( S_\alpha = S_\beta \) and \( t_a = t_\beta \) then a (classical or de se) proposition \( p \) obtains in \( \alpha \) iff it obtains in \( \beta \). Propositions cannot take different truth-values at a pair of cases that differ only with respect to the target proposition singled out.\(^2\) I remain neutral on whether to individuate propositions coarsely as sets of possible worlds or not. We focus on cases not affected by those matters.

A condition obtains or fails to obtain in each case. Conditions are coarsely individuated by the set of cases in which they obtain. The total set of cases includes at least all nomically possible cases. A condition \( C \) entails another condition \( D \) iff in all cases in which \( C \) obtains, \( D \) obtains as well.

Since conditions are coarsely individuated, we need to distinguish them from the definitions, requirements and accounts that express them and state the relations between them.\(^3\) Suppose for instance that some infallible belief analysis of knowledge is true. Then knowing that \( p \) and infallibly believing \( p \) are the same condition: they obtain in the very same cases. Yet the two following accounts of knowledge would still be distinct:

1. \( S_\alpha \) knows that \( p_\alpha \) in \( \alpha \) iff \( S_\alpha \) knows that \( p_\alpha \) in \( \alpha \).
2. \( S_\alpha \) knows that \( p_\alpha \) in \( \alpha \) iff \( S_\alpha \) infallibly believes that \( p_\alpha \) in \( \alpha \).

Similarly, accounts, definitions and requirements have logical structure: they may be conjunctive, or disjunctive and so on. Conditions, by contrast, have no logical structure. We can meaningfully talk of a conjunction of conditions: the conjunction of \( C \) and \( D \) is the condition that obtains in each case in which both \( C \) and \( D \) obtain. But we cannot meaningfully

---

\(^2\) Call a super de se proposition a proposition that takes different value at \( \alpha \) and \( \beta \) while \( \alpha \) and \( \beta \) differ only with respect to the target proposition involved. I suspect that introducing super de se propositions will result in self-referential paradoxes in the definition of cases, unless some typing is introduced to build them.

\(^3\) I normally use “definition” for a definition of a condition, “requirement” for the requirement that a certain condition be satisfied in order for another to obtain, and “account” for a statement of relations between certain condition. But I will not be strict on this. The only important matter is to distinguish the conditions themselves from linguistic/conceptual entities that refer to them.
say that the resulting condition is conjunctive: only the definition of this condition is. This would be like saying that a set is “intersective” because it is identical to the intersection of two sets.

For analogous reasons, we should be wary of saying that conditions on knowledge are modal, defeasibility-based, virtue-theoretic, explanationist, and so on. Definitions, requirements and accounts that refer to such conditions can be couched in modal (proper function, explanationist, etc.) terms or can involve modal (proper function, etc.) concepts. But the condition themselves do not involve terms or concepts. For instance, if the requirement of knowing p and that of infallibly believing p express the same condition, the condition itself is not directly epistemic, modal, or (for that matter) physical. We could be generous and say that a condition is “modal” iff there is some modal requirement or definition that expresses it. But the generous use is trivialisable. Better be strict, and reserve such qualifications to definitions, requirements and accounts themselves.

Our terminology is sometimes at odds with the common usage of “condition” in the epistemology literature (as in “necessary and sufficient conditions for knowledge”). The epistemologists’s use of “condition” is typically ambiguous between a condition and a requirement in the present sense. The ambiguity is often harmless, but not so for our discussion. So we need the slightly more cumbersome language of requirements, definitions and accounts.

### 3.1.2 Fallible and infallible conditions

A condition C is infallible if and only if the target proposition is true wherever C obtains:

**Infallible condition**  C is infallible iff for all cases α, if C obtains in α then $p_α$ is true in α.

---

4. Let C be any condition. Consider the requirement: being in a case α such that necessarily, if a case β is identical to α then C obtains in β. The requirement involves modal concepts. A case satisfies it iff it satisfies C, thus (since conditions are coarse) the condition expressed by the requirement is identical to C. So C is modal.
A condition is fallible iff it is not infallible. Impossible conditions are vacuously infallible; by “infallible condition” we normally mean a non-vacuously infallible one. A requirement is infallibilist iff it requires an infallible condition.

The truth condition of the justified true belief analysis can be defined thus:

**The truth condition**  $T$ is the condition such that for all cases $\alpha$, $T$ obtains in $\alpha$ iff $p_\alpha$ is true in $\alpha$.

It is trivially infallible. It is also the weakest infallible condition: for any condition $C$, if $C$ is infallible then $C$ entails $T$. Thus an infallible condition is a condition that entails the truth condition.

Belief is not infallible:

**The belief condition**  $B$ is the condition such that for all cases $\alpha$, $B$ obtains in $\alpha$ iff one believes $p_\alpha$ in $\alpha$.

For there are some false beliefs, that is, there are some subjects $S$ and propositions $p$ such that $S$ believes $p$ and $p$ is not true. Let $\beta$ be the case constituted by such a pair $S, p$ and the corresponding time and world. $B$ obtains in $\beta$, but $p_\beta$ is not true in $\beta$. So $B$ is not infallibilist.

(Our framework does not allow us to formulate cases of truth-valueless belief. For instance, a subject may believe that *the son of his neighbour is a truck driver*, while his neighbour has two sons and nothing in the context or subject’s intention attaches his thought to one of the sons in particular. One may argue that due to something akin to presupposition failure, the subject’s belief is truth-valueless. The framework does not allow such cases: a case of belief is a case in which the subject believes the distinguished proposition of the case, and we do not allow truth-valueless propositions. For the purpose of the present work, however, we can consider all putative truth-valueless beliefs as being false.)

We can define various justification conditions. For instance, we could assume a three-place justification relation between two propositions and a subject and define:

**The justification condition**  $J$ is the condition such that for all cases $\alpha$, $J$
obtains in $\alpha$ iff in $\alpha$, there is some proposition $q$ such that $q$ justifies $S_\alpha$ in believing $p_\alpha$.

A justification condition $J$ is fallible iff there is some case $\beta$ such that $J$ obtains in $\beta$ and $p_\beta$ is false.

Knowing is of course infallible:

**The knowledge condition** $K$ is the condition such that for all cases $\alpha$, $K$ obtains in $\alpha$ iff one knows $p_\alpha$ in $\alpha$.

For if one knows that $p_\alpha$ in $\alpha$, $p_\alpha$ is true in $\alpha$. The same holds for the conditions expressed by other factive attitude verbs, such as *seeing that*, *learning that*, *proving that*, and so on.

Why not simply talk of “factive” conditions instead of infallible ones? Because I cannot make sense of the difference between “merely factive” and “really infallible” conditions. Suppose Descartes’s story is right: there is a condition $CD$ of clearly and distinctively perceiving the target proposition, such that for any $\alpha$, if $CD$ holds in $\alpha$ then $p_\alpha$ is true. By contrast, consider the condition $TB$ of truly believing such that for any case $\alpha$, $TB$ holds in $\alpha$ iff both $T$ and $B$ obtain in $\alpha$. Descartes’s condition is genuinely infallibilist, the thought goes, while the condition of truly believing is merely factive. Why? I can see two reasons to say this. (1) The condition $TB$ is factive by definition, while it would be a substantial discovery if the condition $CD$ was. (2) The condition $TB$ has a subjective component, belief, which is a fallible mental state. But none of these ideas make sense when talking of *conditions*. Conditions have no definitions; there are no analytic truths about them. Conditions have no components; as we pointed out, a conjunction of conditions like $TB$ is not itself conjunctive. So one cannot sort out conditions into those that are trivially or analytically truth-entailing and those who are substantially so. Given this, we can either refer to those conditions as factive conditions or infallibilist conditions. I opt for the later term because most of the factive/infallible conditions I am interested in would traditionally be described as “infallible”.
A related point is that infallible/factive conditions are necessarily so, on our definitions. A proposition is necessarily true iff it is true in all cases. For any infallible condition C and case α, the proposition that C is infallible is necessarily true if it is true in α. For it is true in α that C is infallible iff the following holds: for any case β, if C obtains in β then \( p_β \) is true in β. Since the latter does not make reference to α, it holds in α iff it holds in any arbitrary case γ. Hence it is necessarily true that C is infallible. Thus the idea of a condition that would be “factive in some cases but not in others” would not make sense.

### 3.1.3 Typology of requirements

Here is how I classify requirements. A requirement states a condition. A condition is infallibilist iff it entails truth; it is fallibilist otherwise. An infallibilist requirement is a requirement that states an infallibilist condition. A fallibilist requirement is a requirement that states a fallibilist condition.

A requirement is modal if it is stated in terms of belief, truth, some notion of basis of belief, and some notion of possibility.\(^5\) By contrast, section 3.2.1 gives examples of infallibilist requirements that are not modal. Modal infallibilist requirements are impossibility of error requirements: they state that some kind of error was in some relevant sense impossible. For short, they are infallibility requirements. (The distinction between “infallibilist requirements,” which may or may not be modal, and “infallibility requirements,” which are modal infallibilist requirements, can be confusing. But from section 3.3 on we focus on infallibility requirements alone. The terminology is only needed for the methodological preamble, and can safely be forgotten afterwards.)

Similarly, fallibilist requirements may or may not be modal. Modal fallibilist requirements require that some kind of error was in some relevant sense almost (hardly, normally, . . . ) impossible. For short, they are reliability requirements.

Modal requirements, fallibilist or infallibilist, have in common the

---

5. Thanks to Pascal Engel here.
requirement that error be avoided in a range of actual and possible cases. I call them *avoidance of error* requirements.

The typology is summarised in table 3.1. In the following section (3.2), I provide methodological reasons to prefer *modal* requirements. This lands us in the bottom row of the table. In the next chapter, I argue that an infallibilist requirement is needed to avoid Gettier problems (4.2). This lands us in the bottom left cell of the table. The rest of this chapter expands that cell by giving the general schema and variants of infallibility requirements (3.3–3.8). (The form is generalised to reliability requirements in 3.6.) In the next chapter, two further constraints narrow down the options to the method infallibilist one I defend.

### 3.2 Motivation for modal requirements

#### 3.2.1 Infallibilist requirements need not be modal

It should be clear by now that infallibilist requirements need not be modal. An infallibilist requirement is a requirement that states an infallibilist condition, i.e. a condition that entails truth. Such requirements can simply state that error is not actual or possible. But they can also state something else which instead entails that error is not actual or possible.
The first ones are modal requirements: they are stated only in terms of belief, truth, some notion of basis of belief, and some notion of possibility. The second are non-modal infallibilist requirements. To illustrate, I give here some examples of requirements that are at least prima facie non modal.

The first example is that of “undefeasibility” or undefeated-justification requirements. Most are infallibilist. For instance:

S knows that \( p \) on the basis of \( j \) only if there is no true proposition \( d \) such that the conjunction of \( d \) and \( j \) does not justify \( S \) in believing \( q \).

The requirement is infallible: for suppose \( p \) is false. Then \( \neg p \) is a defeater of \( S \)’s justification: \( \neg p \) is true and \( \neg p \land j \) does not justify \( S \) in believing \( p \). This particular version of an undefeated-justification was suggested by Chisholm (1964, 147–9) but never endorsed because it was immediately taken to be too strong (see Lehrer and Paxson, 1969, 227–8 and Swain, 1978, 162–4). However, most undefeasibility requirements in the literature are infallibilist (Howard-Snyder et al., 2003, 306).

A second illustration is Plantinga’s (1993, 19) proper function account:

\[ S \text{ knows that } p \text{ iff } S \text{’s belief that } p \text{ is true and based on a module of } S \text{’s cognitive mechanisms that is (a) aimed at truth, (b) functioning properly, (c) functioning in the kind of environment for which it has been designed, (d) reliable, i.e. such that there is a high objective probability that a belief based on that module functioning properly in that kind of environment is true.} \]

In his (1993, 17–9), Plantinga’s requirement (a)–(d) were presented as fallibilist: a belief could satisfy (a)–(d) and be false. But the view faced Gettier problems, as other fallibilist views (Greene and Balmert, 1997). Plantinga (1996, 328; 1997, 144) subsequently added the “resolution condition” that the module should be:

\( (e) \text{ infallible in the mini-environment of } S \text{’s belief that } p \). \]

Conditions (e) cannot be satisfied by a false belief; so the truth condition is redundant and the account is infallibilist. Moreover, condition (e) is a modal infallibilist condition. It is satisfied by a belief case \( \alpha \) iff for all
belief cases whose mini-environment is the same as in $\alpha$, one avoids error based on the same “module”. But Plantinga’s conditions (a)–(d) are not modal. Condition (d) is formulated in terms of probability, not possibility. Conditions (a)–(c) involves notions of design and proper function. As Plantinga notes (1993, 4), these belong to a different family from the modal ones.

Another illustration is that of Sosa’s (2007, 22–3) and Greco’s (2003, 116–7, 123) aptness or credit requirement on knowledge:

$S$ knows that $p$ only if $S$’s belief is true because virtuous.\(^6\)

Trivially, a false belief cannot be true because virtuous. So Sosa’s and Greco’s requirement is also infallibilist. But the requirement is not purely modal. For while Sosa and Greco account for virtue in terms of bases of belief and a modal notion of reliability, their aptness condition makes an essential use of an explanation relation expressed by “because.”\(^7\)

Now it may turn out that some notions that are conceptually distinct from modal notions are nevertheless reducible to modal notions. For instance, one may try to reduce objective probability to some notion of truth in most possible cases. Similarly, one may try to reduce explains to necessary entails. Given some such reduction, a non-modal requirement can be recast as a modal one. But unless some such reduction is appended to them, probabilist, proper function and explanationist accounts are not modal accounts.

Even Unger’s (1968) non-accidentality requirement, which is now widely endorsed as a necessary condition on knowledge, is not clearly a

\(^{6}\) An earlier statement is found in Sosa (1988, 174–5): “We have reached the view that knowledge is true belief out of intellectual virtue, belief that turns out right by reason of the virtue and not just by coincidence.” See also p.174, where the idea of credit appears in connexion with a lucky golf swing case. And in Sosa (1993, 62–3): see the quote and commentary section 3.7.4.

\(^{7}\) A different type of explanationist condition has been put forward by Jenkins (2008, 74–7): $S$ knows that $p$ only if $A$ believes $p$ because $p$ is true. Jenkins’s account can be seen as a generalisation of Alvin Goldman’s (1967) causal account, according to which (very roughly) $S$ knows that $p$ iff $p$ caused $S$ to believe $p$. I am tempted to see the causal account and explanationist accounts as forming a unique family. Other explanationist accounts have been defended by Alan Goldman (1988, 22) and Steven Rieber (1998, 194). See Jenkins (2008, 78–83) for a discussion.
modal requirement:

S knows that \( p \) only if it is not at all an accident that S’s belief that \( p \) is true.\(^8\)

It is tempting to analyse accidentality as relative contingency.\(^9\) Yet it is debatable whether the notion of accidentality is that of relative contingency. Accidentality may rather be improbability, unexplainability, unpredictability or some other notion; it may also be a primitive one not equivalent to any of those. So while Unger’s condition is manifestly infallibilist, it is not obviously modal.

Similar remarks can be made for the widespread idea that knowledge excludes luck (see Pritchard, 2005, 1 and the references in Pritchard, 2005, 126–7). As Pritchard (2005, 127, 130) argues, it is not obvious that the anti-luck and non-accidentality are equivalent. Nor is it evident that luck is a modal notion. (Though one can certainly defend a modal account of it, as Pritchard, 2005, 128 eventually does.)

3.2.2 Methodological considerations for infallibility requirements

If we need an infallibilist condition, we need an infallibilist requirement. But the infallibilist requirement need not be modal: that is, it need not be couched in modal terms of impossibility of error. Here I

\(^8\) For sample endorsements see Sosa (1988, 184), Sosa (1997b, 417–9), Greco (2003, 116), Williamson (2000, 268), though Williamson prefers to keep “accidental” in quotation marks.

\(^9\) Unger’s notion cannot be analysed as simple contingency. Unger (1968, 160) allows that (1) it is not at all accidental that \( p \) (my belief is true), (2) it is accidental that \( q \) (I am alive), (3) \( p \) entails \( q \). If “it is not at all accidental that” is taken as “it is necessary that”, we have: (1) \( \Box p \), (2) \( \neg \Box q \), (3) \( p \rightarrow q \) is valid. This is incompatible with orthodox (i.e. normal) modal logics. By Necessitation and (3), we have \( \Box (p \rightarrow q) \). By Distributivity (axiom K), we have \( \Box p \rightarrow \Box q \), so either (1) or (2) is false. Yet we can still analyse accidentality in terms of relative contingency. Say that it is not at all accidental that \( p \) iff \( \Box(c \rightarrow p) \), where \( c \) specifies initial conditions relevant to \( p \). We then get (1*) \( \Box(c \rightarrow p) \) and (2*) \( \neg \Box(c' \rightarrow q) \). From (1*) and (3) it follows that \( \Box(c \rightarrow q) \), i.e., the relevant initial conditions for \( p \) include the fact that I am alice, and from this and (2*) it follows that \( \Diamond(c' \land \neg c) \), so the relevant initial conditions for \( p \) and for \( q \) are distinct, and \( \Diamond \neg c \), so \( c \) is contingent. (Revised. Thanks to Timothy Williamson.)
3.2. Motivation for modal requirements

present two methodological reasons to (at least initially) try to account for knowledge in modal terms, on the assumption that we need an infallibility condition (as the historical perspective of chapter 2 suggests, and as I will argue in the next chapter, 4.2).

The first reason is that infallibility requirements (that is, impossibility of error requirements) are the most straightforward way to get an infallibility condition. We need a requirement that implies that error is avoided: why not simply require that error be avoided? Infallibility requirements are requirements that simply state that one is in an infallible condition. They are a straightforward way to satisfy the need for an infallibility condition.

The second reason is that modal notions are currently better understood than competing concepts. The last fifty years or so have given us good logics, models and concepts for modalities. Modal conditions are thus relatively easy to formalise and their consequences are more readily seen. The same cannot be said about notions of justification and explanation.

The Bayesian tradition provides rigorous probabilistic characterisations of justification. 10 But it is difficult to get an infallibility condition out of those. 11 Other substantial theories of justification are not easy

10. de Finetti (1989); Jeffrey (1965, 1983); see also Ramsey (Dokic and Engel, 2001, §1.2 and §2.1).

11. For even probability 1 does not entail truth. The chance that a point-like dart randomly thrown on a continuous line will not fall on a given point x is 1, even though it is not impossible that the arrow fails on x. See Williamson (2007a).

Revision note. Igor Douven asked why this weaker notion of infallibility (a method is infallible iff the probability that its outputs are true is 1) is not sufficient for knowledge. It is difficult to build intuitive cases where probability 1 and necessity fall apart. Here is an attempt: if you only know that the clock is going to ring in the next minute, you cannot know of any particular instant that the clock will not ring at that instant, even if it will not in fact ring at that instant, and even though the probability that it rings at that instant is 0. But one may object to the case that we are not even able to refer to point-like instants, which is sufficient to explain why we cannot know anything about it, or that if we use devices to refer to instants (e.g. by talking about the exact middle of the minute), it is not so clear that we do not know that the clock will not ring at that very instant (provided it does not). A somewhat weaker motivation for the infallibility approach vs. the probability-1 approach is that the later requires adding truth as a separate condition on knowledge, while the infallibilist approach gives a simple characterization.

Igor Douven further pointed out that there are alternatives to the probability-1 option within the Bayesian approach. One could build on the notion of resilience
Many epistemologists are happy to follow Chisholm (1977, 6–7) in using a primitive notion of justification while providing even less characterisation of it that Chisholm did. For instance, Conee and Feldman (2004a, 102) rely on a primitive notion of a doxastic attitude “being on balance supported by a person’s evidence.” One is certainly entitled to have primitive notions. But that does not preclude answering such questions as whether evidential support is closed under logical implication, whether a visual experience evidentially supports all propositions entailed by its content, whether a visual experience evidentially supports the proposition that one has that experience, what the structure of evidential “strength” is like, how thresholds for evidence to be “strong enough” behave (does one threshold fit all or are they proposition-dependent or time-dependent or stake-dependent and so on?). Answering such questions would give us a better grasp of the notion.

That many philosophers do not feel the pressure to do so despite the last thirty years of unresolved disputes over justification is surprising, to say the least. There is no clearly dominant framework for the notion of explanation developed by Skyrms and others: we do not merely require that (say) the probability that a method outputs a true belief is 1, but also that it is one conditional on most or all propositions that are consistent with the method outputting a belief. This indeed an interesting path to explore, though it remain to be seen whether such strengthened probabilist requirements will not turn out to be substantially equivalent to the modal infallibilist formulation.

12. See Pollock (2008) for some references and a proposal.
13. Even Chisholm’s (1977, 22, 24) notions of the “self-presenting” and the “directly evident” are ultimately defined in terms of the primitive notion of an doxastic attitude being more reasonable than another for a given subject, proposition and time (and, presumably, a world).
14. Conee and Feldman (2008) specify that evidence is individuated internally and that evidence supports the hypothesis that best explains its existence. But that does not yet settle any of the questions above. A prima facie dilemma for their view is that they hold both that (1) what one is justified in believing is only a function of the evidence one has and (2) what hypothesis best explains one’s evidence is relative to the hypotheses that a subject can think of. Taken together the claims implies the mere (in)ability to entertain some hypotheses is part of one’s evidence. Either they have to broaden the notion of evidence so that one’s conceptual repertoire is part of it, or to give up one of (1) or (2).
3.2. Motivation for modal requirements

Some accounts are themselves framed in probabilistic or modal terms, so they may not provide a genuine alternative. Moreover, a number of theories of explanations focus on causal explanation and are consequently not readily applicable to epistemology. For neither the condition that \( p \) causally explains \( S \)'s belief that \( p \) nor the condition that \( S \)'s belief that \( p \) causally explains \( p \) can be accepted as necessary or sufficient for knowledge without a lot of fine-tuning or — depending on how one sees it — epicycles.\(^{16}\) Hence an explanationist account has to say what notion of explanation is used, and it may have to clarify the notion itself (as Greco, 2003, 132 admits). None of this, of course, means that explanationist accounts are mistaken, even less that we should not look for an account of the relevant notion(s) of explanation. But it seems to me preferable to try modal accounts first. It may well turn out that explanationist and modal accounts are not in conflict, if the relevant explanationist turns out to be reducible to modal ones.

This is rash, of course. I do not deny that there are substantial and worthwhile accounts of justification and of explanation, nor do I claim that all modal conditions are clearer and better understood than any justificationist or explanationist condition. My point is only that overall, modal notions are (currently) simpler to use and understand, especially for epistemological purposes. This provides a very defeasible consideration in favour of trying a modal account first.\(^{17}\)

---

15. See Woodward (2003, esp. chap 8) for an overview and proposal.
16. See Goldman (1967, 364–6, 368–70) for an illustration.
17. Revision note. Ernest Sosa pointed out that the argument of this section is misleading. Some of the modal requirements I discuss, and in particular the one I defend, appeal to a notion of method or basis of belief. But the concept of method, if cashed out in terms of cognitive setups (which may involve mental states and processes, properties of the environment, and so on) and some epistemically relevant dimension of similarity across cognitive setups, is no less obscure than the concepts of justification, aptness or proper function.

While the point is well taken with respect to the argument above, the formal model of methods in part III is meant to alleviate this kind of worry. The formalism fixes structural features of methods. For instance, methods are individuated by their conditional and unconditional outputs; the way in which one gets a given output does not matter. This gives us a grasp of the notion which is partial: for instance, the formal characterization alone does not help us to the concept to a particular case. But a grasp nevertheless: we
3.3 The avoidance of error

I am interested in requirements on knowledge formulated in terms of truth, belief, some notion of bases of belief, and some notion of possibility. Call them *modal requirements*, for short. More precisely, my interest is with requirements according to which knowledge excludes certain actual and possible errors. Call these *avoidance of error* requirements. I provide here a general form for such requirements. Such requirements can be presented as necessary, sufficient or necessary and sufficient for knowledge. Many of the modal accounts of knowledge put forward in the last decades can be put into this form.

3.3.1 Error, ignorance, falsity

Recall that a *case* is a centred world with a distinguished proposition (3.1.1). A *condition* obtains or fail to obtain in each case. Belief and truth are conditions:

**The truth condition** For any case $\alpha$, $T\alpha$ iff $p_\alpha$ is true in $\alpha$.

**The belief condition** For any case $\alpha$, $B\alpha$ iff $S_\alpha$ believes $p_\alpha$ in $\alpha$.

Where $p_\alpha$ and $S_\alpha$ are the distinguished proposition and subject of case $\alpha$.

We can now introduce the conditions of *error* and *ignorance*. We will say that one is in *error* if one believes a falsehood, and that one is in *ignorance* if one is ignorant of a truth. We will say that one is in *falsity* if one is false and does not believe it.

Ernest Sosa makes a second objection to the methodological case for the modal approach. Ultimately, I argue that our understanding of the notion of method should be guided by our intuition about knowledge and the method-infallibilist account of knowledge (see section 4.3). Why would alternative accounts not use the same strategy? If they did, the present argument against them fails: while notions of justification, proper function or aptness may not be *prima facie* well understood, nothing has been said to show that our understanding of them will not improve when they are embedded in an account of knowledge.

I grant the point. The only thing that I can say at this stage is that with few exceptions, alternative accounts have detailed enough so as to allow one to build formal models in which their implications can be explicitly drawn. By contrast, many modal accounts are relatively straightforward to formalize — the present chapter offers a step in that direction —, including the one defended here, as shown in Part III.
3.3. The avoidance of error

ignorance if one fails to believe a truth. (My use of “ignorance” may depart from ordinary usage; but we need a convenient name for the absence of belief in a truth.) Error and ignorance are the two elementary mismatches between belief and fact:

**Error** For any $\alpha$, $E\alpha$ iff $B\alpha \land \neg T\alpha$.

**Ignorance** For any $\alpha$, $E\alpha$ iff $T\alpha \land \neg B\alpha$.

*Infallibility* is a maximum in avoidance of error. We may call *omniscience* a maximum in avoidance of ignorance.\(^{18}\) Most of the requirements that will occupy us are avoidance of error requirements.

(Recall that for our purposes truthvalueless beliefs, if there are any, count as false: see p. 106 above.)

The notion of error can be extended in two directions. First, a modal condition may not merely require avoidance of error but avoidance of *falsity* at relevant cases (Williamson, 2009a, 23, see below 3.4.6). Thus we may introduce falsity alongside error and ignorance:

**Falsity** For any $\alpha$, $F\alpha$ iff $\neg T\alpha$.

Interestingly, requiring the avoidance of falsity or the presence of truth is equivalent — if bivalence holds. Thus Falsity can be equally seen as an extension of the notion of ignorance.

Second, some modal conditions require the avoidance of some specific form of error — typically, the avoidance of error on some specified basis. We will introduce the relevant notation as we proceed.

### 3.3.2 Overview of modal requirements

A modal requirement states a condition that obtains in a target case $\alpha$ as a function of whether there are errors of a certain type in other possible or actual cases $\beta$, $\gamma$, . . . . Modal requirements are built out of four components:

**Grounding condition (Ga)** A condition that the target case must satisfy.

18. I owe this use of the infallibility/omniscience distinction to Egré (2008). This use of course diverges from the standard notion of omniscience as knowledge of everything.
Respect of similarity \((\alpha R \beta)\) A reflexive accessibility relation between cases that specifies in which cases \(\beta\) one is to avoid error in order to satisfy the condition in \(\alpha\). Typically: that the case \(\beta\) involves the same subject, the same proposition, and probably the same time, and that it be “close” to \(\alpha\) is some relevant sense.\(^{19}\)

Error type \((E\alpha, E_m\alpha)\) The type of error to be avoided. Typically false belief: \(B\alpha \land \neg T\alpha\).

Proportion \((A,M)\) How much error should be avoided in the range. We’ll consider only two variants: in all cases (infallibility) and in most (reliability).

Simple infallibility requirements straightforwardly implement the idea (3.4). Parametric infallibility requirements add a parameter which allows a coordination between the grounding condition, the respect of similarity and the error type (3.5). Reliability requirements generalise to avoidance of error in most cases (3.6). Double requirements combine two avoidance of error requirements, typically an infallibility one and a reliability one: the grounding condition and error condition is the same, but there are two respects of similarity and typically two distinct proportions for each (3.7). All schemas can be seen as special cases of the double requirement one. My preferred view, however, is to see the parametric one as the most fundamental, and double requirements as a conjunction of two parametric requirements.

Most modal accounts in the literature fit into one of the schemas. Seeing them in this light brings to light a number of structural properties of interest. Let me quickly point them out here. We return to them in the summary of the results (3.8).

A requirement is proposition-centred, subject-centred and time-centred iff its range is restricted to cases involving the same proposition, subject and

\(^{19}\) To be a proper similarity relation, \(R\) needs to be symmetric as well as reflexive. Symmetry will clearly fail for some requirements, e.g. Nozick’s. For others it will clearly hold, e.g. Armstrong’s. I have not investigated such questions in detail. Note also that \(R\) is not an epistemic accessibility relation (a relation \(E\) such that if \(K\alpha\) and \(\alpha E\beta\) then \(p_\alpha\) is true in \(\beta\)); for this reason, the symmetry of \(R\) is not straightforwardly linked to the validity of some sceptical arguments (compare Williamson, 2000, 164–7).
3.3. The avoidance of error

time, respectively:

**proposition-centring** \( R \) is such that \( p_\beta = p_\alpha \) for any \( \beta \) such that \( \alpha \R \beta \).

**subject-centring** \( R \) is such that \( S_\beta = S_\alpha \) for any \( \beta \) such that \( \alpha \R \beta \).

**time-centring** \( R \) is such that \( t_\beta = t_\alpha \) for any \( \beta \) such that \( \alpha \R \beta \).

Proposition-centred requirements fail to satisfy what I call the Method Constraint (4.1). Subject-centring is discussed in appendix E. Time-centring raises interesting issues, but I will not discuss it further in the present work.\(^{20}\)

Other features of interest are *variability* and *fallibility*. Fallibility has already been defined (3.1.3). Fallibilist requirements fail the Infallibility Constraint (4.2).

Given a modal requirement with a certain relation \( R \), and a case \( \alpha \), we can define the set of *worlds* in which the relevant alternative cases to \( \alpha \) are located:

**relevant worlds** The set \( \mathcal{R}_\alpha^R \) of relevant worlds for \( \alpha \) in view of \( R \) is the set of all worlds \( w_\beta \) such that \( \alpha \R \beta \) for some \( \beta \).

\[
\mathcal{R}_\alpha^R = \{w : \exists \beta (\alpha \R \beta \land w = w_\beta)\}.
\]

I call a requirement *variable* iff its sets of relevant worlds \( \mathcal{R}_\alpha^R \) do not merely depend on the world, subject and time of \( \alpha \), but also on the target proposition \( p_\alpha \). I call a requirement *fixed* iff its set of relevant worlds is the same for any two cases involving the same subject, world and time. Variable requirements tend to violate deductive closure. Appendix D illustrates the point with time-relative safety.

A synoptic presentation of the results of this review is provided in tables 3.2–3.4 (p. 147–149). The notation introduced as we go along is summed up in the List of Symbols (E.3, p. 459).

\(^{20}\) The discussion in appendix D concerns variability due to *time-indexed closeness*, not to time-centring itself.
3.4 Simple modal requirements

3.4.1 The schema

In the simpler requirements, the grounding condition is just belief \((B\alpha)\), and the error to be avoided is just false belief \((B\beta \land \neg T\beta)\). We have:

**Simple infallibilist requirement schema** Where \( R \) a reflexive relation over cases, define \( IE \) such that for any case \( \alpha \), \( IE\alpha \) iff:

\[
B\alpha, \text{ and }
\forall \beta \text{ such that } \alpha R \beta \text{ , } \neg E\beta.
\]

\[
B\alpha \land \forall \beta (a R \beta \rightarrow \neg E\beta).
\]

It is easy to see that the requirement entails true belief. Suppose \( IE \) holds in \( \alpha \). So we have \( B\alpha \). Since \( R \) is reflexive, \( aRa \), and since \( IE \) holds in \( \alpha \), \( \neg Ea \), that is \( \neg (B\alpha \lor \neg Ta) \). Since \( B\alpha \), \( Ta \). Such requirements are only satisfied in cases of true belief. Thus they entail truth and are infallibilist in our sense.

3.4.2 True belief and infallible belief

Some simple illustrations first. The true belief requirement \( TB \) is obtained with \( R \) defined as follows:

**True belief (TB)** For any \( \alpha, \beta, a R \beta \text{ iff } \alpha = \beta. \)

\( TB \) requires the target case to be a case of belief and not to be a case of error. It is easy to see that it is satisfied iff one has a true belief in the target proposition.

For infallible belief — a requirement sometimes tentatively attributed to Descartes —, we define:

**Infallible belief (IB)** For any \( \alpha, \beta, a R \beta \text{ iff } B\beta \text{ and } p_{\beta} = p_{\alpha}. \)

\( IB \) requires that the subject believes the proposition in \( \alpha \), and that error is avoided in in all cases of belief in the same proposition. (Including cases

21. Williamson’s revised safety requires avoidance of falsity instead (3.4.6).

22. As proved earlier, \( TB\alpha \) entails true belief: \( Ta \land Ba \). Conversely, suppose that \( Ta \) and \( Ba \) hold in some case \( \alpha \). Then \( B\alpha \). Let \( \beta \) be any case such that \( a R \beta \). By the definition of \( R, \beta = \alpha \). Since \( Ta \) holds in \( \alpha \), \( T\beta \). So \( \neg E\beta \). So \( T Ba \).
where the belief is held by a different subject.) Some *de se* propositions like “I exist” or “I have a belief” satisfy it.

### 3.4.3 Goldman’s relevant-alternatives condition and Lewis’s infallibilist contextualism

Goldman’s (1976, 778–84) relevant-alternatives condition for perceptual knowledge is a restricted form of the infallible belief condition. Goldman introduces two weakenings of infallible belief. First, one should avoid error only at *relevant* alternative cases, not all possible cases (1976, 774–8). Second, one should only avoid error in cases that are *perceptually equivalent* to the target case, not all relevant cases. Roughly, a case $\beta$ is perceptually equivalent to a target case $\alpha$ iff the subject in $\alpha$ cannot distinguish being in $\alpha$ from being in $\beta$ (1976, 778–84). We define:

**Goldman’s relevant alternatives (GRA)** $\alpha R \beta$ iff $p_\beta = p_\alpha$, $\beta$ is a relevant alternative to $\alpha$ and $\beta$ is perceptually equivalent to $\alpha$.

Goldman’s characterisation of perceptual equivalence makes clear that a case perceptually equivalent to $\alpha$ must involve a (living) subject positioned in a certain way relative to certain objects and undergoing some perceptual experience — at least as long as $\alpha$ is itself a case involving some perceptual experience. It is not clear whether the subject has to be the same as in the target case; so we leave this aspect open here. Note also that the relevant-alternative relation is a relation between *cases* and not between *worlds*. That would allow the requirement to exclude some *actual* cases of errors about the same proposition as being nevertheless irrelevant. But Goldman does not discuss such issues.

**Lewis** (1996, 551–3) proposes a contextualist variant of Goldman’s relevant-alternative condition. (The idea was earlier put forward by

---

23. Wherever the subject of $\alpha$ is not existent, not alive or without experience, we assume that any case is equivalent to it. This ensures that $R$ is reflexive. This also implies that sleeping subjects do not have perceptual knowledge, but the consequence is acceptable — they have only mnemonic knowledge, which is not covered by the account.

24. Like the closeness relation in Williamson’s safety, see 3.4.6 below.
Stine (1976) and Lewis (1979) and partly anticipated by Goldman (1976, 776–7). In Lewis’s version, each context fixes a set of relevant alternative possibilities. So we need to relativise the requirement to conversational contexts. Perceptual equivalence is formulated more narrowly as having the same experiences and memory as in actuality. (It is also unclear whether Lewis requires the subject to be different across perceptually equivalent cases: I’ll leave it open here as well.  

25

Lewis’s contextualist infallibilism (LCI) For any context $c$ and case $\alpha$, $\alpha$ satisfies LCI relatively to $c$ iff:

$\text{Ba and } \neg E \beta$ for any $\beta$ such that $\alpha R \beta$, where

$\alpha R \beta$ iff $\beta$ is a member of the set of alternatives for context $c$, and $S_\beta$ has the same experiences and memories in $\beta$ as $S_\alpha$ in $\alpha$.

Note though that on this formulation, the relations $R$ are typically not reflexive. On pain of trivialisation, some contexts will count some cases as irrelevant. Let $\gamma$ be a case irrelevant in $c$. By the definition of $R$, $\neg \gamma R \gamma$. So $R$ is not reflexive. This has the disastrous consequence that factivity fails in Lewis’s account, as we argue in ch. 6 (sec. 6.6.2).

3.4.4 Nozick’s sensitivity and tracking

The simple version of Nozick’s truth-tracking account (1981, 172–7) requires true belief, sensitivity and adherence, where the latter are characterised as follows, for a subject $S$ and a proposition $p$:

Sensitivity If $p$ were not true, $S$ would not believe that $p$.  

Adherence If $p$ were true, $S$ would believe that $p$.  

25. See his formulations in Lewis (1996, 552–3), for instance: “I say that the uneliminated possibilities are those in which the subject’s entire experience and memory are just as they actually are.” (553) The natural reading is that the subject of the possibility in question is the same as the actual one (since “they” binds “the subject’s entire experience”). On the other hand, Lewis evokes the possibility that one does not know who he is (p. 555). He may have realised that on his account that is possible only if relevant alternative cases includes ones in which someone else has the same evidence.

26. The sensitivity requirement was earlier suggested by Dretske (1971, 1) and Armstrong (1973, 169).

27. Nozick’s actual requirement is “if $p$ were true, $S$ would believe $p$ and would not believe $\neg p$.” I ignore the complication.
3.4. Simple modal requirements

(Dretske (1971, 1–8) had suggested a similar view.28)

Let us write \( p \rightarrow q \) for the subjunctive conditional “if \( p \) were true, \( q \) would be true.” Nozick adopts a variant of the standard Stalnaker-Lewis semantics for subjunctive conditionals (Stalnaker, 1968; Lewis, 1973). The semantics assume a three place relation \( x \sqsubseteq y \ z \) over worlds such that for any \( w \), \( \sqsubseteq_w \) is a complete preorder over a set of worlds including \( w \).29 When \( w_1 \sqsubseteq_w w_2 \) holds we say that \( w_1 \) is at least as close to \( w \) as \( w_2 \). When \( w' \) is out of the order of \( w \), \( w' \sqsubseteq_w w' \) fails and we say that \( w_1 \) is inaccessible from \( w \). \( w \) is assumed to be as close to itself as any world in the order: for any \( w' \) such that \( w' \sqsubseteq w \), \( w \sqsubseteq w' \). Each world \( w' \) in the order of \( w \) determines a sphere of worlds that are at least all close to \( w \) at \( w' \).

On the standard Stalnaker-Lewis semantics,30 the subjunctive conditional \( p \rightarrow q \) holds at \( w \) iff there is no accessible \( p \)-world, or there is some accessible \( p \)-world \( w' \) such that \( p \rightarrow q \) holds throughout the sphere of worlds closer to \( w \) than \( w' \) (Lewis, 1973, 16). We say that a semantics of subjunctive conditionals is variable iff the sphere of worlds through with \( p \rightarrow q \) has to be true varies with the antecedent proposition. The Stalnaker-Lewis semantics is variable for conditionals with false antecedents. Where \( p \) and \( p' \) are both false, \( p \rightarrow q \) requires the truth of \( p \rightarrow q \) up to some \( p \)-world, while \( p' \rightarrow q' \) requires the truth of \( p' \rightarrow q' \) up to some \( p' \)-world.

The Limit assumption is the idea that for any world \( w \) and antecedent \( p \), there is a set of \( p \)-worlds that are the closest to \( w \) (Lewis, 1973, 19-20). That is, there are worlds \( w' \) such that \( p \) holds at \( w' \) and for any world \( w'' \)

28. For all Dretske says in his (1971), he could have had a fixed-threshold semantics for subjunctive conditionals that would result in a safety-like condition; in particular, we could formalise the requirement in Dretske (1971, 1–2) roughly as Sosa’s reliable indication safety (see below 3.5.4). It is only Dretske’s (1970) endorsement of what are in fact consequences of the sensitivity requirement — the failure of deductive closure and the related diagnosis of sceptical arguments — that justify the attribution of a sensitivity-like view to him. Curiously, the alternative-based semantics for subjunctive conditionals that he sketches in his (1970) does not appear in his (1971). I do not attempt to formalise Dretske’s account here.

29. Preorders allow ties but no incomparabilities. See appendix A for details.

30. Revision note. This paragraph and the next have been revised. Thanks to Timothy Williamson for pointing out to me that original formulations relied on the Limit assumption.
in the order of \( w \), if \( p \) holds at \( w'' \) then \( w' \preceq w'' \). If the limit assumption holds, the semantics is equivalent to saying that \( p \rightarrow q \) holds at \( w \) iff \( q \) holds at all \( p \)-worlds closest to \( w \). Though the assumption is dubious (Lewis, 1973, 20), it will simplify our discussion to make it. As far as I can see, the assumption does not affect our conclusions.

With this semantics and the limit assumption, sensitivity is an infallibilist requirement characterised as follows:

**Sensitivity (SE)** \( \alpha R \beta \) iff \( p_\beta = p_\alpha \), \( S_\alpha = S_\beta \), and \( w_\beta \) is at least as close to \( w_\alpha \) as any \( \neg p_\alpha \) world.\(^{31}\)

\( \text{SE} \) requires belief in \( p_\alpha \) in the actual world, and that the subject avoids false belief that \( p_\alpha \) in a “sphere” of cases whose worlds are at least as close as the closest \( \neg p_\alpha \) worlds. Since \( w_\alpha \) is maximally close to itself, \( \alpha \) is included in the sphere, and so \( p_\alpha \) must be true in \( \alpha \). Now consider any other case \( \beta \) such \( \alpha R \beta \). If \( w_\beta \) is a world in which \( p_\alpha \) is true, then \( \beta \) cannot be an error case, since \( p_\beta = p_\alpha \) and \( p_\alpha \) is true. (Thus \( \text{SE} \) does not require belief in the intervening \( p_\alpha \) worlds that are closer than the first \( \neg p_\alpha \) ones.)

The crucial case is when \( w_\beta \) is one of the closest not-\( p_\alpha \) worlds. Since \( p_\beta = p_\alpha \) and \( \neg p_\alpha \) in \( w_\beta \), \( \beta \) is a case of error iff it is a case of belief. Thus the requirement is satisfied only if the subject does not believe \( p_\alpha \) in the closest worlds in which it is false.\(^{32}\)

If we assume, as Lewis (1973, 14) does, that each \( w \) is the *unique* world closest to itself, a subjunctive conditional with a true antecedent is true if its consequent is true. This would trivialise Nozick’s adherence condition. A stronger truth condition for conditionals with truth antecedents is needed. There are several options, which we review in appendix A. Nozick’s own proposal is that when \( p \) is true, \( p \rightarrow q \) is true iff \( q \) holds at all \( p \)-worlds up to but excluding the first \( \neg p \) ones (Nozick, 1981, 680n).

For simplicity, I will use the slightly stronger quasi-Nozickian semantics

\[^{31}\text{On Nozick’s view, } \alpha R \beta \text{ probably also requires that } t_\alpha = t_\beta, \text{ or some such restriction. Presumably, whether if } p \text{ had been false I would have wrongly believed it to be true in the distant past or future does not matter to whether I know } p \text{ now.}\]

\[^{32}\text{SE is not strictly equivalent to Nozick’s third condition, since it additionally implies true belief (Nozick’s conditions 1 and 2). But that is all the better, I think, and at any rate harmless.}\]
below:

**Quasi-Nozickian semantics** $p \square \rightarrow q$ is true at $w$ iff $p \rightarrow q$ holds throughout a sphere $S_{w,p}$, where $S_{w,p}$ extends up to and including the closest worlds to $w$ where $p$ differs in truth-value from $w$.

That is, for any $w', w' \in S_{w,p}$ iff $w'$ is at least as close to $w$ than all worlds $w''$ where $p$ holds if and only if $p$ does not hold in $w$: $w' \in S_{w,p}$ iff $\forall w''((w'' \models p \leftrightarrow w \not\models p) \rightarrow w' \preceq_w w'')$.

The quasi-Nozickian semantics requires that $q$ be true in every $p$-world (a) up to the closest $p$-worlds if $p$ is false, and (b) up to the closest $\neg p$-world if $p$ is true. The semantics is equivalent to the Lewis-Stalnaker one when the antecedent is false. The semantics is variable for both conditionals with a false antecedent and conditionals with a true one.

The adherence condition is an impossibility of ignorance requirement. We formulate it by replacing the no-error condition with a no-ignorance condition. With quasi-Nozickian semantics, the result is:

**Quasi-Nozickian Adherence (QNA)** For all $\alpha$, $QNA\alpha$ iff $Ba$ and $\neg I\beta$ for all $\beta$ such that $\alpha R\beta$, where:

$\alpha R\beta$ iff $p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is at least as close to $w_\alpha$ as any $\neg p_\alpha$ world.

Recall that $I\alpha$ iff $T\alpha \land \neg Ba$. Since it is required that the subject believes in $\alpha$, $\neg I\beta$ is trivially the case in $\alpha$ — whether or not $p_\alpha$ is false in $\alpha$, so the condition does not entail truth. The condition requires that the subject believes $p_\alpha$ in all the worlds where $p_\alpha$ is true up to the closest $\neg p_\alpha$ ones.

Quasi-Nozickian Adherence and Sensitivity are easily unified. Taken together, adherence and sensitivity require avoidance of both error and ignorance across a single sphere of worlds, namely all the worlds up to and including the first $\neg p$ ones:

---

33. The semantics is stronger than Nozick’s for conditionals with true antecedents, but weaker for conditionals with false antecedents. See appendix A for details.

34. Williamson (2000, 149) calls “Nozickian” a semantics that uses a fixed threshold of close worlds instead of Lewis’s comparative closeness. This is not Nozick’s actual semantics, but rather what we call a *hybrid* semantics. See appendix A.
Quasi-Nozickian tracking (QNT) For all $\alpha$, $QNT\alpha$ iff $Ba$ and $\neg E\beta \wedge \neg I\beta$ for all $\beta$ such that $\alpha R\beta$, where $\alpha R\beta$ iff $p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is at least as close to $w_\alpha$ as any $\neg p_\alpha$ world.

Equivalently: $NT\alpha$ iff $Ba$ and $B\beta \leftrightarrow T\beta$ for all $\beta$ such that $\alpha R\beta$.

The requirement entails true belief at the actual world. It is satisfied when one avoids ignorance and error that $p$ up to including the closest worlds in which $p$ is false. That is one construal of Nozick’s account of knowledge. Nozick’s actual semantics is a bit more complicated, and the resulting account slightly different, but we leave out the details in the appendix A.

3.4.5 Sosa’s safety and tracking

Variable version

Sosa (1996, 274–6, 1999b, 142, 2007, 25) contrasts Nozick’s sensitivity with his safety requirement. In its simple version, Sosa’s safety is characterised as follows:

**Sosa’s Safety** If $S$ believed $p$, $p$ would be true.

Sosa prefers the formulation “$S$ would believe $p$ only if $p$ were true,” which is less odd to use when it is true that $S$ believes $p$, as it is in the cases of interest. 35 But Sosa takes the “only if” conditional to be equivalent to the “if” one given above. 36 Sosa originally maintained Nozick’s adherence

---

35. Another formulation, which Sosa credits Carl Ginet for, is “It would not be so that $S$ believes that $p$ without it being so that $p$” (Sosa, 2002, 285). This seems to me dangerously close to “$S$ would not believe $p$ without $p$ being true,” which in turn seems to be a slight variant of “$S$ would not believe $p$ if $p$ was not true,” namely, Nozick’s sensitivity conditional.

36. The equivalence between “$p$ only if $q$” and “if $p$, $q$” seems right with indicative conditionals. (“Alice passed only if she worked”; “If Alice passed, she worked’’.) Not so with subjunctive ones: “would $p$ only if $q$” appears to differ from “if $q$ would $p$.” (“Sophie would see Pedro dance only if she went to the Parade”; “If Sophie saw Pedro dance, she would go to the Parade.”) But this may be superficial; perhaps the two phrases trigger different pragmatic effects for a same logical form, in the way “They got married and had children” and “They had children and got married” do.

Here is a test case. Two similar glass balls, $A$ and $B$, are about to be dropped, and I am asked whether I think they will break. I do not know, but I think that they are
3.4. Simple modal requirements

condition ("counter-safety" in his 2000a, 40), and called the resulting requirement Cartesian tracking (1996, 276–7, 1999a, 373, 2000a, 38, 2002, 265). But he later dropped the adherence condition (1999a, 376, 2000a, 40–1, 2002, 274). 37

Sosa’s safety, like Nozick’s adherence, is trivialised on semantics like Lewis’s on which the truth of both $p$ and $q$ entails that of $p \rightarrow q$, since any true belief would then satisfy it. Let us first interpret it with the simple Quasi-Nozickian semantics. 38 The result is:

**Variable Safety (VS)** $\text{VS} \alpha$ iff $B \alpha$ and $\neg E \beta$ for all $\beta$ such that $\alpha R \beta$, where $\alpha R \beta$ iff $p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is as close to $w_\alpha$ as any world in which $S$ fails to believe $p_\alpha$.

Variable safety requires one to avoid error up to the closest worlds in which one *fails to believe* $p_\alpha$, instead of requiring avoidance of error up to the closest worlds in which $p_\alpha$ is *false*. (On the Nozickian semantics, it would require avoidance of error up to but excluding such worlds; on the quasi-Nozickian semantics, it requires error up to and including them.)

Cartesian tracking requires Adherence as well as Safety. With Quasi-Nozickian Adherence defined as previously, we get the following unified definition of Cartesian tracking:

---

37. The reasons to drop it are (1) glimpse cases, in which one glimpses by chance e.g. a bird passing by the window: one knows despite one’s belief not being “adherent”, and (2) it was only needed because sensitivity is trivially satisfied by beliefs in necessary truths. Nozick (1981, 179, 193) deals with glimpse cases by introducing methods, see below 3.5.3.

38. Sosa (1996, 276n) does not endorse a particular semantics, but only points out that he rejects the “usual” semantics on which true-true conditionals are true. The main hints he provides for the truth condition of $p \rightarrow q$ conditionals with true antecedent are (1) as the situation in fact stands, $p$ would be true only if $q$ would be true (Sosa, 1996, 275), (2) as a matter of fact, though perhaps not as a matter of strict necessity, $p$ would true without $q$ being true (1999a, 378), (2) Not easily would $p$ be true without $q$ being true, given the actual setup (1999a, 382, 2002, 279). (See also Sainsbury, 1997, 907–8 on the latter formulation.)
Variable Cartesian Tracking (VCT) \( VTC\alpha \) iff \( Ba \) and \( \neg E\beta \land \neg I\beta \) for all \( \beta \) such that \( aR\beta \), where:

\[ aR\beta \text{ iff } p_\beta = p_\alpha, S_\alpha = S_\beta, \text{ and } w_\beta \text{ is at least as close to } w_\alpha \text{ as any world} \]

where \( S \) fails to believe \( p \), or at least as close to \( w_\alpha \) as any world where \( p_\alpha \) is false.

Equivalently: \( VCT\alpha \) iff \( Ba \) and \( B\beta \leftrightarrow T\beta \) for all \( \beta \) such that \( aR\beta \).

The rough idea is this. Imagine the worlds standing in a line extending from the actual world. Safety requires the truth of \( p \) up to the first non-belief world. Adherence requires belief up to the first non-truth world. If truth “stops” before belief does, Safety is violated. If belief “stops” before truth does, Adherence is violated. Together, both conditions require a belief/truth match up to the first non-belief worlds, or the first not-\( p \) worlds, depending on which one is the most distant. The only way to satisfy both is to have the first non-belief world to be also the first non-truth world.\(^{39}\)

We get this surprising result: Cartesian tracking is stronger than Quasi-Nozickian tracking. Roughly put, Quasi-Nozickian tracking requires avoid-

---

\(^{39}\) Proof of \( VS\alpha \land QNA\alpha \leftrightarrow VCT\alpha \). (I will talk of cases being closer, rather than of their respective worlds being closer. The proper formulation is easily recovered but cumbersome.)

Left-to-right. Suppose \( VS \) and QNA hold in \( \alpha \). Assume for reductio that VCT is false in \( \alpha \). \( Ba \) holds, by VS. So it must be that \( E\beta \lor I\beta \) for some \( \beta \) such that \( aR\beta \), with \( R \) defined as in VCT. By that definition, either (1) \( \beta \) is at least as close to \( \alpha \) as any world where \( S \) does not believe \( p \), or (2) it is at least as close to \( \alpha \) as any not-\( p \) world. Suppose (1). Since VS holds, \( \neg E\beta \). So it must be that \( \beta \). Now since QNA holds, \( \neg I\gamma \) for any \( \gamma \) that is as close to \( w_\alpha \) as any \( \neg p_\alpha \)-world; so \( \beta \) cannot be as close to \( w_\alpha \) as any \( \neg p_\alpha \)-world. Thus let \( \gamma \) be some \( \neg p_\alpha \)-world closer than \( \beta \). Since \( \beta \) is closer to any non-belief world and \( \gamma \) is closer than \( \beta \), it follows that \( \gamma \) is a belief world. So we have \( By \land \neg Ty \), that is \( E\gamma \). But since \( \beta \) is closer than any non-belief world and \( \gamma \) is closer than \( \beta \), \( VS\alpha \) is violated, contrary to assumption. Now suppose (2). Since QNA holds, \( \neg I\beta \). So it must be that \( E\beta \). Since \( VS\alpha \) holds, \( \beta \) cannot be as close to \( \alpha \) as any non-belief world. So let \( \gamma \) be some non-belief world closer than \( \beta \). Since \( \beta \) is at least as close as any not-\( p \) world, and \( \gamma \) is closer than \( \beta \), \( \gamma \) is a \( p \)-world. So \( Ty \land \neg B\gamma \), that is \( I\gamma \). But since \( \beta \) is closer than any not-\( p \) world, and \( \gamma \) is closer than \( \beta \), \( \gamma \) is closer than any not-\( p \) world. Thus \( VA\alpha \) is violated, contrary to assumption. So if \( VS\alpha \land QNA\alpha \), then VCT\( \alpha \).

Right-to-left. Suppose \( VCT\alpha \). Then \( Ba \). Let \( \beta \) be any case such that \( p_\alpha = p_\beta, S_\alpha = S_\beta \), and \( \beta \) is as close to \( \alpha \) as any non-belief world. Since VCT\( \alpha \), \( \neg E\alpha \). So \( VS\alpha \) holds. Now let \( \gamma \) be any case such that \( p_\gamma = p_\alpha, S_\gamma = S_\alpha \), and \( \gamma \) is as close to \( \alpha \) as any \( \neg p \)-world. Since VCT\( \alpha \), \( \neg I\alpha \). So QNA\( \alpha \) holds.
3.4. Simple modal requirements

ance of error and ignorance that $p$ up to the closest $\neg p$ worlds. But Cartesian tracking requires avoidance of error and ignorance that $p$ up to the closest $\neg p$-worlds and up to the closest non-belief worlds. This is surprising, because Cartesian tracking is supposed not to have the sceptical consequences that Nozickian tracking has (Sosa, 1996, 277). Yet it has sceptical consequences in virtue of the adherence condition that have been overlooked because it was assumed that the sceptical consequences of Nozick’s view came entirely from its sensitivity condition.

Take, for instance, the belief that it is not the case that I am a brain in vat ($\neg h$). The closest worlds in which I am one are supposed to be “far out”. Now consider all the worlds that are strictly closer than them. By Adherence, one has to believe $\neg h$ in all those closer worlds. If the brain-in-a-vat world is indeed “far out”, this is a difficult requirement to satisfy. For instance, there may be some situation in which some philosopher would persuade me not to believe that I am not a brain in a vat. If that situation counts as strictly closer to actuality than the brain in a vat worlds, I fail to satisfy Adherence. (The point holds whether we construe Adherence on the quasi-Nozickian semantics or on the Nozickian semantics.) So the Adherence condition has sceptical consequences as well, under the variable semantics.

Another perhaps surprising result is that Variable Safety violates deductive closure. Deductive closure is the idea that if one knows that $p$ and competently deduces that $q$, then one knows that $q$. Nozick’s sensitivity is notoriously incompatible with closure (Nozick, 1981, 204–11). Sosa’s Safety was meant to avoid this consequence (Sosa, 1996, 277). But it does not, as Sosa (1999a, 378, 1999b, 149) acknowledges.

41. Nozick (1981, 179) avoids the extreme sceptical consequences of adherence by introducing methods.
42. Sosa does not explain why he takes the simple version of his Safety to violate closure. I suspect that he had in mind the second source of incompatibility between simple safety and closure, namely that simple safety is not relativised to bases or methods — see the discussion of fixed safety below. That would be consistent with the fact that he takes the lack of closure to motivate the introduction of a reference to bases of belief (1999a, 378–9) instead of a modification of the semantics for conditionals.
knows that $p$ and competently deduces $q$ from $p$. Variable Safety implies that $S$ avoids error that $p$ up to the closest worlds in which $S$ fails to believe $p$. But the Variable Safety requirement on the conclusion requires avoidance of error that $q$ up to the closest worlds in which $S$ fails to believe $q$. The first does not include the second unless $S$ would not believe $q$ unless she believed $p$. But it is wrong to assume that whenever $S$ believes $q$ on the basis of deduction from $p$, $S$ would not believe $q$ if she did not believe $p$.

**Fixed version**

Williamson (2000, 149) offers (but does not endorse) a semantics for the Safety and Adherence conditionals that avoids both sceptical consequences for Adherence and (one source of) violation of closure for Safety. On this semantics, when $p$ is true, the conditional $p \Box \rightarrow q$ is true iff $p \rightarrow q$ holds at close worlds. The set of close worlds is the same whichever antecedent $p$ is considered. Thus I call the truth condition Williamson suggests for conditionals with true antecedent a fixed-threshold one. It contrasts with the variable one of quasi-Nozickian semantics. 43

The resulting requirements are:

**Fixed Safety (SA)** $SA\alpha$ iff $Ba$ and $\neg E\beta$ for any $\beta$ such that $p_\beta = p_\alpha$, $S_\alpha = S_\beta$ and $w_\beta$ is close to $w_\alpha$.

**Fixed Adherence (FA)** $FA\alpha$ iff $Ba$ and $\neg I\beta$ for any $\beta$ such that $p_\beta = p_\alpha$, $S_\alpha = S_\beta$ and $w_\beta$ is close to $w_\alpha$.

**Fixed Cartesian tracking (FCT)** $FCT\alpha$ iff $Ba$ and $\neg E\beta \land \neg I\beta$ for any $\beta$ such that $p_\beta = p_\alpha$, $S_\alpha = S_\beta$ and $w_\beta$ is close to $w_\alpha$.

Equivalently, iff $Ba$ and $Ba \leftrightarrow Ta$ for any such $\beta$.

Most authors assume a fixed-threshold semantics for the safety requirement (Sainsbury, 1997, 913–4; Williamson, 2000, 127–8; Vogel, 2000, 604, 2007, 83; Pritchard, 2002b, 297, 2005, 71; see also Heller, 1995, 506 and 1999b, 200–1 for a similar idea without the “safety” label). 44

43. See Appendix A for details. Williamson’s semantics is hybrid, since the truth conditions for conditionals with false antecedents remains variable.

44. Pritchard’s (2002b, 297) “super-safety” is equivalent to (the fixed version of) Sosa’s Cartesian tracking.
Even on the fixed-threshold semantics, the simple safety requirement is not compatible with deductive closure (Sosa 1999a, 378, 1999b, 149). S may know that p and competently infer that q, and still there may be a close world in which S would have a false belief that q on some other basis than deduction from p.\footnote{Revision note. To expand a bit: this is simply because simple safety is not relativized to bases. Grandma sees her grandson and sees that he is sick. She infers (rightly) that one of her grandchildren is sick. But had she not seen him, she would have been convinced that he was sick, irrespectively of whether he was and against all assurances to the contrary. In some such possibility she infers the false conclusion that one of her grandchildren is sick. She does not satisfy simple safety, as formulated, though she clearly knows that he is sick.}

Moreover, the simple safety requirement faces a number of apparent counterexamples. Take Goldman’s (1976, 779) Dachshund case, for instance:

S sees a Dachshund in the field and forms the belief that there is a dog there. But in the circumstances there could easily have been a wolf instead of the dog, and S would have mistaken the wolf for a dog.

With appropriate filling in, S knows that there is a dog in front of her, even though there is a close world in which she believes that wrongly. Intuitively, the reason why the close error is nevertheless irrelevant is that the close error would have been made on a significantly different basis. Both considerations motivate the formulation of safety in reference to bases of belief (Williamson, 2000, 128; Sosa 1999a, 378, 1999b, 149; 2005, 156). This requires us to refine our general schema for infallibility requirements (3.5).

Let me mention in passing a strengthening of Safety that has sometimes been considered. The idea is that one should not only avoid (close) errors that p but also errors about p, that is, one should both avoid falsely believing that p and falsely believing \( \neg p \).\footnote{See Becker (2006, 698) for the proposal and Hiller and Neta (2007, 311–2) for a criticism.} The idea straightforwardly results from:

**Safety about** \( p \alpha R \beta \) iff \( p_\beta = p_\alpha \) or \( p_\beta = \neg p_\alpha, S_\alpha = S_\beta \), and \( w_\beta \) is close to \( w_\alpha \).
3.4.6 Williamson’s safety

Williamson’s (2000, 100, 147) safety requirement is formulated differently:

**Williamson’s Safety (WS)** For any $\alpha$, $WS\alpha$ iff $B\alpha$ and $\neg E\beta$ for any $\alpha R\beta$

where:

$aR\beta$ iff $S_\beta = S_\alpha$ and $\beta$ is sufficiently similar to $\alpha$.  

Williamson suggests at times that (WS) should be qualified, for instance by adding that error should only be avoided in sufficiently similar cases in which one’s belief has a similar basis (2000, 102, 147; 2009c, 325). But that is not necessary. One can instead leave open what constitutes relevant similarity. Bases will play an important role, but other aspects will do to, such as the level of confidence (2000, 98–9) or the proposition believed (2000, 102, 2009c, 325), and it is not necessary nor perhaps possible to give a full account of them. Most importantly, it is doubtful that relevant similarity can be factored into several respects of similarity, as Williamson sometimes suggests (see ch. 4, sec. 4.1.3).

Note that the requirement is very open. A relevant case $\beta$ need not involve the same time, nor the same proposition (2000, 102, 2009c, 325). In fact, we could also drop the constraint that the case involve the same subject. This allows Williamson’s safety to avoid some of the problems that beset proposition-centred accounts (see again sec. 4.1.3).

As stated, safety still requires that one avoids error, that is, false belief at close cases. Williamson has recently suggested an amendment to require the avoidance of falsity instead (2009a, 23):

**Williamson’s Revised Safety (WRS)** For any $\alpha$ $NWS\alpha$ iff $B\alpha$ and $\neg F\beta$ for any $\beta$ sufficiently relevantly similar to $\alpha$.  

47. Williamson’s formulations are ambiguous on whether the subject must be the same in the relevant error cases. He seems to assume it, since he often allows the in principle case-relative pronoun “one” to scope over cases operators. For instance: “If one believes $p$ truly in a case $\alpha$, in which other cases must one avoid false belief in order to count as reliable enough to know $p$ in $\alpha$?” (Williamson, 2000, 100). I take up the question whether relevant errors should be subject-centred in appendix E.

48. Cf. “The suggestion is that knowing $p$ requires safety from the falsity of $p$ and of its epistemic counterparts.” (Williamson, 2009a, 23) and the model that Williamson
For instance, if one believes that a certain coin will land heads, the actual case may be similar enough to a case constituted by: a world in which the coin lands tails and the proposition that it lands heads. Actuality is, so to speak, too close to the situation of the proposition *the coin lands heads* in some world in which it lands tails. On that view, a close situation need not involve a subject or a belief for it to make the actual belief unsafe.

3.5 Parametric modal requirements

3.5.1 The schema

A number of modal requirements, including the one I defend, can only be formulated in a refined version of the general schema:

**Parametric infallibility requirement schema** Where $G_i$ and $E_i$ are conditions with parameter $i$ and $R_i$ a reflexive relation over cases with parameter $i$, define $IE$ such that for any case $\alpha$, $IE\alpha$ iff:

For some $i$, $G_i\alpha$ and for all cases $\beta$ such that $\alpha R_i \beta$, $\neg E_i \beta$.

$$\exists i(G_i\alpha \land \forall \beta(\alpha R_i \beta \rightarrow \neg E_i \beta)).$$

The grounding condition $G_i\alpha$ and the type of error $E_i\alpha$ are now represented by two-place predicates with $i$ a parameter and $\alpha$ a case. They must be a belief condition and an error condition, respectively:

**Grounding condition** For any $i, \alpha$, $G_i\alpha \rightarrow B\alpha$.

**Error type** For any $i, \alpha$, $E_i\alpha \rightarrow E\alpha$.

As stated, the requirement does not entail true belief. But that will be so for all the particular values of $G$ and $E$ we will consider. That is, $G$ and $E$ are always such that:

$$(G_i\alpha \land \neg E_i\alpha) \rightarrow T\alpha.$$
The introduction of a parameter allows us to coordinate some aspect of the grounding condition with the type of error and with the respect of similarity. In fact, in all but one of the accounts we consider, the respect of similarity $R_i$ does not depend on the value of $i$; coordination between the grounding condition and the error type is sufficient. So we will almost always use a simple non-parametric relation $R$.

For a toy illustration, consider the requirement of an infallibly based belief.

**Infallibly based belief (IBB)** $E_i\alpha$ iff in $\alpha$, $S_\alpha$ believes that $p_\alpha$ on basis $i$ and $p_\alpha$ is false.
$G_i\alpha$ iff in $\alpha$, $S_\alpha$ believes that $p_\alpha$ on basis $i$.
$aR_\beta$ iff $p_\alpha = p_\beta$.

The requirement is satisfied in $\alpha$ iff $S_\alpha$ believes that $p_\alpha$ on some basis such that in no case does one believes $p_\alpha$ on that basis while $p_\alpha$ is false. It is not required that any belief that $p_\alpha$ be true; only that any belief on that basis be true. Note that $(G_i\alpha \land \neg E_i\alpha) \rightarrow T\alpha$ holds for any $\alpha$, as required.

The requirement IBB involves coordination of grounding and error type. We could in principle get an equivalent requirement by coordinating the grounding condition with the respect of similarity instead:

$E_i\alpha$ iff $E\alpha$,
$G_i\alpha$ iff in $\alpha$, $S$ believes that $p_\alpha$ on basis $i$.
$aR_i\beta$ iff $p_\alpha = p_\beta$ and in $\beta$, $S_\beta$ believes that $p_\alpha$ on basis $i$.

On this formulation, we restrict the range of cases to cases in which the proposition is believed on the same basis. We can then simply require the avoidance of straightforward error.

However, I find that the formulations that “pack” more into the similarity relation rather than into the error type are less clear. One reason for this is that we lose the reflexivity of $R$. For instance, with the formulation above, we get for a given $i$ and an arbitrary case $\gamma$: $\gamma R_i \gamma$ iff $p_\gamma = p_\gamma$ and in $\gamma$, $S_\gamma$ believes that $p_\gamma$ on basis $i$. So if $\gamma$ is not a case of belief, we do not have $\gamma R \gamma$. Another, non principled, reason is that it simplifies comparison with the method infallibilist requirement I favour. Thus wherever possible, I
opt for the formulations that keep $R$ reflexive and non-parametric.

### 3.5.2 Armstrong’s nomological infallibilism

On Armstrong’s (1968, 189, 1973, 168) account, non-inferential knowledge consists in a “law-like connection” between belief and truth. If we construe “law-like connection” as nomic necessity, his account is a parametric infallibilist requirement.\(^{49}\) In its simpler version:

**Armstrong’s simple nomological account** \(E_C\alpha \iff C\alpha \land E\alpha,\)
- \(G_C\alpha \iff B\alpha \land C\alpha,\)
- \(αRβ \iff p_β = p_α\) and the laws of $α$ hold in $β$,
- where $C$ is a relevantly general specification of a subject’s conditions and circumstances.

$R$ is reflexive and \(G_C\alpha \land \neg E_C\alpha \iff T\alpha\), so the requirement entails true belief. It further requires avoidance of error in any nomically possible case of belief in the same proposition in which $C$ holds. This is a variant of the infallibly based belief condition, where $C$ plays the role of the basis.

On Armstrong’s (1973, 170, 182, 197) more refined statement, non-inferential knowledge does not merely require a law-like connection between believing $Fa$ and $Fa$ for some particular proposition $Fa$ (where $F$ is a predicate and $a$ a singular term), but a more general law-like connection between believing $Fx$ and $Fx$ for any $x$. This revised requirement fits our schema as follows. First, assume that propositions are structured. Second, write $p[a/b]$ for the proposition that results from substituting the singular term/concept for $a$ in $p$ by the term for $b$ in all its occurrences. \(G_C\alpha\) and \(E_C\alpha\) are defined as before, and $R$ is modified to:

**Armstrong’s nomological account** \(αRβ \iff \exists x\exists y(p_β = p_α[x/y])\) and the laws of $α$ hold in $β$.

The requirement is only meant to apply to knowledge of singular propositions (1973, 192–4).

\(^{49}\) Thanks to Pascal Engel here.
3.5.3 Nozick’s sensitivity and tracking with methods

Nozick amends his tracking account by considering the methods with which a belief is formed (Nozick, 1981, 179). If in actuality $S$ believes that $p$ via method $m$, the new requirements are:

**Nozick’s sensitivity with methods** If $p$ was not true and $S$ were to use $m$ to arrive at a belief whether (or not) $p$, then $S$ would not believe, via $m$, that $p$.

**Nozick’s adherence with methods** If $p$ was true and $S$ were to use $m$ to arrive at a belief whether (or not) $p$, then $S$ would not believe, via $m$, that $p$.

We give the Nozickian formalisation in Appendix A. Here we give the quasi-Nozickian one. We abbreviate:

$Bm\alpha$ in $\alpha$, $S_\alpha$ believes that $p_\alpha$ on the basis of $m$.

We define:

**Quasi-Nozickian M-sensitivity (QNT)** $E_{m\alpha}$ iff $Bm\alpha \land \neg T\alpha$.

$G_{m\alpha}$ iff $Bm\alpha$.

$\alpha R_{m\beta}$ iff $p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is at least as close to $w_\alpha$ as any world where $p_\alpha$ is false and $S_\alpha$ uses $m$ to arrive at a belief whether (or not) $p_\alpha$.

**Quasi-Nozickian M-adherence (QNT)** $I_{m\alpha}$ iff $Bm\alpha \land \neg T\alpha$.

$G_{m\alpha}$ iff $Bm\alpha$.

$\alpha R'_{m\beta}$ iff $p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is at least as close to $w_\alpha$ as any world where $p_\alpha$ is false or $S_\alpha$ fails to use $m$ to arrive at a belief whether (or not) $p_\alpha$.  

---

50. Here is how the definitions are obtained. Write $Bm?p$ for “$S$ use $m$ to arrive at a belief whether (or not) $p$”. For a given case $\alpha$, the antecedent of the first conditional is $\neg p_\alpha \land Bm?p_\alpha$, the antecedent of the second is $p_\alpha \land Bm?p_\alpha$. The conditionals are applied in cases $\alpha$ such that $Bm\alpha$. (In other cases, the requirements are not satisfied since $G_{m\alpha}$ fails). Consider first those such that $T\alpha$, that is, $p_\alpha$ holds. The antecedent of the first conditional is false, the antecedent of the second is true. On the Quasi-Nozickian semantics, a subjunctive conditional is true iff the corresponding material conditional is true up to and including the closest world where the antecedent has a different truth value. This leads us up to worlds such that $\neg p_\alpha \land Bm?p_\alpha$ for the first conditional, but to worlds such
3.5. Parametric modal requirements

The two requirements cannot be unified as in the simple version (sec. 3.4.4). Method-relative Adherence and Sensitivity do not operate on the same sphere: on the method-relative version of Adherence, it is not required that one believes \( p \) up to the closest worlds in which \( p \) is false, but only up to the closest worlds in which one “ceases” to use the method actually used. This is how method-relative Adherence avoids the sceptical consequences of simple Adherence (see sec. 3.4.5 on the latter).

**Quasi-Nozickian M-tracking (NMT)** For any \( \alpha \), \( \text{QNMT}_\alpha \) iff there is an \( m \) such that \( Bm\alpha \) and:
- \( \neg I_m\beta \) for any \( \beta \) such that \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \), and \( w_\beta \) is at least as close to \( w_\alpha \) as any world where \( p_\alpha \) is false or \( S_\alpha \) uses \( m \) to arrive at a belief whether (or not) \( p_\alpha \).
- \( \neg E_m\beta \) for any \( \beta \) such that \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \), and \( w_\beta \) is at least as close to \( w_\alpha \) as any world where \( p_\alpha \) is false and \( S_\alpha \) uses \( m \) to arrive at a belief whether (or not) \( p_\alpha \).

Nozick’s M-tracking is the only requirement for which we cannot avoid introducing a parameter on the respect of similarity relation \( R \). Not only the error to be avoided has to be an error made by using the same method, but the range of worlds in which error has to be avoided also depends on the method used in the target case.

Luper-Foy (1984, 28–9) and Williamson (2000, 154) point out problems arising from the fact that methods are mentioned in the antecedent of the sensitivity conditional and thereby affect the range of relevant worlds. They suggest that Nozick should drop the reference to methods in the antecedent. (Neither of them endorse the resulting view.) The

---

51. Suppose methods are finely individuated so as to require some form of “perceptual equivalence” (as in Goldman, 1976). Suppose I am currently seeing a tree. What are the closest worlds in which there is no tree, but I use a similar perceptual experience to decide whether there is a tree? Presumably worlds in which I hallucinate a tree, and it is likely that I wrongly believe that there is a tree in some such world. So even in ordinary cases, the sensitivity conditionals would take one to “far out” worlds and would hardly ever be satisfied. But on the other hand, Nozick would need finely individuated methods to deal with cases like Goldman’s Dachshund (cf. 131 above).
amended sensitivity requirement is just like Nozick’s original one but for the relativisation of the relevant type of error and ignorance to methods. Assuming $S$ believes that $p$ via $m$:

**Revised Nozick’s sensitivity with methods** If $p$ was not true, then $S$ would not believe, via $m$, that $p$.

**Revised-Nozickian M-sensitivity (RNMS)** For any $\alpha$, $RNMS\alpha$ iff there is an $m$ such that $Bm\alpha$ and $\neg E_m\beta$ for any $\beta$ such that $\alpha R\beta$, where:

- $E_m\alpha$ iff $Bm\alpha \land \neg T\alpha$, and
- $\alpha R\beta$ iff $p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is at least as close to $w_\alpha$ as any $\neg p_\alpha$ world.

The respect of similarity relation need not be parametric here.

An analogous revision for Adherence is inadvisable. If no reference to methods is made in the antecedent of Adherence, then Adherence requires one to believe up to the closest $\neg p$ world, which may be a “far out” world depending on $p$. We would be saddled again with the sceptical consequences of adherence (sec. 3.4.5).

### 3.5.4 Basis-relative safety

Several proponents of safety want to relativise safety to bases (Williamson, 2000, 128; Sosa 1999a, 378; 1999b, 149; 2005, 156; Pritchard, 2005, 152–5). As before, we abbreviate:

- $Bmp$ iff $S$ believes that $p$ on the basis of $m$.

The fixed version of basis-relative safety is given by:

**Basis-relative safety (BS)** $E_m\alpha$ iff $Bm\alpha \land \neg T\alpha$

- $G_m\alpha$ iff $Bm\alpha$,
- $aR\beta$ iff $p_\beta = p_\alpha$, $S_\alpha = S_\beta$, $w_\beta$ is close to $w_\alpha$.

That is, one believes on a basis such that one avoids false belief on the same basis at close worlds. We assume a fixed threshold for closeness, but of course a variant with a variable threshold could be formulated. This version of Safety, fixed threshold included, has been endorsed by Pritchard (2005, 162). (Pritchard (2010) has added a proposition-uncentred virtue
3.6. Reliability requirements

requirement and has recently (p.c.) endorsed a proposition-uncentred safety requirement closer to method infallibilism.) It is the requirement that is most commonly called “safety” in the literature. 52

Sosa’s (1999a, 378–9) preferred way to introduce bases was through a reliable indication theory. (Sosa (2007, 30) has since given up safety in favour of aptness.) An indication is an internal state of mind that in some sense “says” to the subject that a certain proposition is true. 53 In our framework, we can describe indications as conditions on cases:

\[ I_\alpha \text{ iff in } \alpha, S_\alpha \text{ has an indication } I \text{ of } p_\alpha. \]

In a given case, a subject’s belief can be based (or not) on a given indication. We write:

\[ B_{I_\alpha} \text{ iff in } \alpha, S_\alpha \text{ believes that } p_\alpha \text{ on the basis of } I. \]

One believes on the basis of \( I \) only if \( I \) obtains: for any \( I, \alpha, B_{I_\alpha} \rightarrow I_\alpha. \)

Now Sosa’s requirement is given by:

Sosa’s reliable indication \( E_{I_\alpha} \text{ iff } I_\alpha \land \neg T_\alpha, \)

\[ G_{I_\alpha} \text{ iff } B_{I_\alpha}. \]

\[ \alpha R \beta \text{ iff } p_\beta = p_\alpha, w_\beta \text{ is close to } w_\alpha. \]

On this version, the indication itself must be sufficient for truth of the proposition at close cases, not merely belief based on the indication. Sosa presented the requirement as necessary but not sufficient for knowledge (1999a, 383n8). Sosa (2002, 267–76) develops further the idea of reliable indication, but I am not sure how to formalise the developments.

3.6 Reliability requirements

3.6.1 The schema generalised

Up to now we have only considered infallibility requirements, that is, requirements that ban error in all cases of a certain range and type.

52. See e.g. Comesaña (2005, 397), who mistakenly attributes it to Williamson (403n4).
53. I gather that indications are internal from the qualifier “ostensible” (Sosa, 1999a, 378), which Sosa (1985, 228) normally uses for the internal counterpart of perceiving, remembering, intuiting and so on.
But infallibility requirements are a special case of the general form of avoidance of error requirements, which ban error in a certain amount of cases of a certain range and type. We can thus generalise our schema as follows:

**Avoidance of error schema** Where $G_i$ and $E_i$ are conditions with parameter $i$ and $R_i$ a reflexive relation over cases with parameter $i$, define $AE$ such that for any case $\alpha$, $AE\alpha$ iff:

For some $i$, $G_i\alpha$ and for all/most/almost all/... cases $\beta$ such that $\alpha R_i \beta$, $\neg E_i \beta$.

$\exists i(G_i \alpha \land [Q : \alpha R_i \beta][\neg E_i \beta])$, where $Q$ is a generalised quantifier.

Assuming a measure on the set of cases, we can define various quantifiers $Q$ such that $[Q : A][B]$ holds iff a certain relation between $A$ cases and $C$ cases obtains. For our present purposes, we need only use a most-like quantifier $M$ such that $[M : A][B]$ iff there is a high ratio of $A \land B$ cases among $A$ cases.

Avoidance of error requirements with a quantifier $Q$ such that $[Q : A][B]$ does not entail $[All : A][B]$ are fallibilist. For such quantifiers, the truth of $[Q : Ax][Bx]$ is compatible with $Ax \land \neg Bx$ for some $x$. In our case, this means that $AE\alpha$ may hold for $i$ even though $\alpha R_i \beta \land E_i \beta$ holds. In particular, $AE\alpha$ may hold even though $\alpha R_i \alpha \land E_i \alpha$: that is, the target case may itself be a case of error. Leaving aside uninteresting versions with quantifiers like “no”, “almost none”, “very few” and so on — the corresponding requirements would be presence of error requirements —, we may call the resulting requirements reliability requirements.

### 3.6.2 Goldman’s process reliabilism

In its simple version, Goldman’s (1979, 13) process reliabilism is the requirement given by:

---

54. See also section 3.1.3 above.
55. Generalised quantifiers were introduced by Lewis (1970a) and independently by Montague (1974).
3.6. Reliability requirements

**Reliabilism** \( E_m\alpha \text{ iff } Bm\alpha \land \neg T\alpha. \)

\( G_m\alpha \text{ iff } Bm\alpha. \)

\( aR\beta \text{ iff } \text{ and } w_\beta \text{ is close to } w_\alpha. \)

Where \( m \) ranges over cognitive processes (Goldman, 1979, 13) and worlds close to actuality are “relevant counterfactual situations, i.e. ones that are highly ‘realistic’ or conform closely to the circumstances of the actual world” (Goldman, 1979, 13). \( Q \) is a most-like quantifier. To spell it out:

**Goldman’s Reliability (GR)** For any \( \alpha \), \( GR\alpha \text{ iff } Bm\alpha \text{ and } \neg (Bm\beta \land \neg T\beta) \)

for most \( \beta \) such that \( w_\beta \) is close to \( w_\alpha. \)

Note that the reliability requirement does not keep the subject or the proposition fixed, by contrast with Nozick’s sensitivity or Sosa’s safety, for instance. Errors made by other subjects about other propositions than \( p \) are relevant to one’s reliably believing that \( p \), provided that those errors are made on the basis of the same process that lead to one’s belief that \( p \).

Goldman considers two interesting variants of the requirement, aimed at dealing with the kind of problem that Sosa (1991, 132) has called the New Evil Demon problem. The first (Goldman, 1979, 17) is given by:

**Goldman’s Rigidified Reliability (GRR)** \( aR\beta \text{ iff } \beta \text{ is close to } @, \)

where @ is the actual world.

On this version, reliability is indexed to the ratio of true beliefs in worlds close to the actual one. (Thus the Evil Demon victim is justified because her belief-forming processes are reliable in worlds like ours, even though they are not reliable in worlds like hers).

The second (Goldman, 1986, 107; anticipated in Goldman, 1979, 18) is given by:

---

56. As Goldman (1979, 10) suggests, instead of treating reliability as a categorical property, we can naturally define degrees of reliability, corresponding to a scale of strength for “most”.

57. In a nutshell, the problem is this: GR is supposed to be necessary and sufficient for justified belief. But it does not seem necessary, since the victim of an Evil demon is justified by processes that are presumably not reliable in worlds like hers (Cohen, 1984). And it does not seem sufficient, since a wishful thinker living in a world where wishful thinking is reliable would still not be justified (Goldman, 1979, 16).

58. The “rigidity” terminology appears in Goldman (1986, 107).
Goldman’s Normalised Reliability (GNR) \( \alpha R \beta \text{ iff } w_\beta \text{ is a normal world,} \)
where a normal world is a world that is consistent with our (actual) general beliefs about the actual world.

For instance, we think that our world is one in which visual perception (from close enough, in good lightning conditions) is reliable. Thus a case of belief satisfies GNR iff it is based on vision, independently of whether (a) vision is reliable in that case’s world, and (b) vision is indeed reliable in the actual world.

On both the rigidified and normalised versions of reliabilism, the relation \( R \) is not reflexive anymore. (Hence it cannot be regarded as a similarity relation.) Reliability in \( \alpha \) requires avoidance of error in (most of) a privileged set of worlds which may not include \( \alpha \). The reflexivity of \( R \) is meant to exclude actual error in the case of infallibility requirements. Since process reliabilism is a fallibilist requirement anyway, this may be tolerable. But still the feature is problematic in an account of knowledge (as opposed to justification): if a subject has a true belief based on a process reliable in normal environments but totally unreliable in hers, she does not know. So at the very least rigidified or normalised reliability with true belief is not sufficient for knowledge.

### 3.6.3 Pritchard’s most-worlds safety

Pritchard’s (2005, 156; 2007, 281) most-nearby-worlds safety is also a reliability requirement. (In view of the lottery problem, Pritchard (2005, 163; 2007, 283) opted instead for an infallibilist requirement.) We get it with:

**Pritchard’s most-worlds safety (PMS)** \( E_{m \alpha} \text{ iff } Bm \alpha \land \neg T \alpha. \)

\(~\)

\( G_{m \alpha} \text{ iff } Bm \alpha. \)

\( \alpha R \beta \text{ iff } p_\beta = p_\alpha, S_\beta = S_\alpha, \text{ and } w_\beta \text{ is close to } w_\alpha. \)

The only structural difference between Goldman’s reliability and Pritchard’s most-world safety is that Pritchard’s requirement only concerns errors involving the same subject and proposition. If we were dealing with all-
3.7 Double modal requirements  

A number of authors have put forward double requirements of avoidance of error, according to which knowledge require both infallibility (in some respect) and reliability (in some other respect). Typically the two requirements are coordinated, so double requirements are not reducible to a simple conjunction of a reliability condition and an infallibility one.

3.7.1 The schema

**Double requirement schema** Where \( G_m \) and \( E_m \) are conditions with parameter \( m \) and \( R^I, R^R \) two reflexive relations over cases, define \( DR \) such that for any case \( \alpha \), \( DR\alpha \) iff:

For some \( i \), \( G_i\alpha \) and

for all cases \( \beta \) such that \( \alpha R^I\beta, \neg E_m\beta \),

for most cases \( \gamma \) such that \( \alpha R^R\gamma, \neg E_m\gamma \).

\[ \exists m (G_m\alpha \land \forall \beta(\alpha R^I\beta \rightarrow \neg E_m\beta) \land [M : \alpha R^R\gamma][\neg E_m\gamma]) \]

3.7.2 Goldman’s global and local reliability

Goldman (1986, 47) says that knowledge requires both “global” and “local” reliability, where global reliability concerns all the outputs of a process, while local reliability concerns its reliability with respect to a

---

59. We leave the parametric \( R \) case aside.
particular proposition. He refers to his (1976) for local reliability (see 3.4.3 above) and his (1979) for global reliability (see 3.6.2 above). We thus have:

**Goldman’s double reliability (GDR)** \( E_m \alpha \iff Bm \alpha \land \neg T \alpha \),

\( G_m \alpha \iff Bm \alpha \),

\( aR^1 \beta \iff p_\beta = p_\alpha, \beta \) is a relevant alternative to \( \alpha \) and \( \beta \) is perceptually equivalent to \( \alpha \).

\( aR^2 \beta \iff w_\beta \) is close to \( w_\alpha \).

It is not clear to me whether Goldman intended to identify the relevant alternative relation and the closeness relation. I leave that open here.

### 3.7.3 Pritchard’s double safety account

Pritchard’s (2007, 292) revision of his safety principle is a double requirement. In his (2005, 156, 161–5) Pritchard considered a fallibilist safety requirement (3.6.3) but opted for an infallibilist one in view of lottery cases. (See also Pritchard, 2007, 281–4.) Greco (2007, 299-301) pressed the dilemma that infallibilist safety was excessively sceptical and fallibilist safety too permissive (see ch. 4, sec. 4.2.1). In his reply, Pritchard (2007, 292) puts forward a double requirement that relies on a distinction between close and very close worlds:

**Pritchard’s double safety (PDS)** \( E_m \alpha \iff Bm \alpha \land \neg T \alpha \),

\( G_m \alpha \iff Bm \alpha \),

\( aR^1 \beta \iff p_\beta = p_\alpha, S_\beta = S_\alpha, w_\beta \) is very close to \( w_\alpha \),

\( aR^2 \beta \iff p_\beta = p_\alpha, S_\beta = S_\alpha, w_\beta \) is close to \( w_\alpha \).

The requirement is still proposition-centred, and thus fails to satisfy the Method constraint (ch. 4, sec. 4.1). Pritchard (2010) replaces proposition-centred most-worlds safety with an uncentred virtue constraint, and has recently adopted an uncentred (p.c.) version of all-worlds safety.

### 3.7.4 Sosa’s and safety

Sosa (1993) suggests a dual requirement close to Goldman’s:
If a faculty operates to give one a belief, and thereby a piece of direct knowledge, one must have some awareness of one’s belief and its source, and of the virtue of that source both in general and in the specific instance. Hence it must be that in the circumstances one would (most likely) believe P iff P were the case — i.e., one (at least probabilistically) tracks the truth (which is part of what is involved in the source’s operating virtuously in the specific instance). And that must be so, moreover, because P is in a field of propositions F and one is in conditions C with respect to P, such that believing a proposition in field F, while one is in conditions C with respect to it, would make one very likely to be right. And, finally, one must grasp that one’s belief non-accidentally reflects the truth of P through the exercise of such a virtue. This account therefore combines requirements of tracking and nonaccidentality, of reliable virtues or faculties, and of epistemic perspective. (Sosa, 1993, 62–3)

The main thrust of the passage is on the need for an epistemic perspective, that is, a set of second-order beliefs of the agents about his epistemic faculties. The epistemic perspective dimension is specific to what Sosa elsewhere calls “reflective knowledge” as opposed to “animal knowledge” (1985, 241–2, 1997b, 422, 2001, 193–4, 2004, 2007, 24). We cannot integrate this dimension into our requirements at this stage.

However, the two other requirements mentioned by Sosa, tracking and reliable virtue, can be taken to constitute an account of animal knowledge. Now Sosa’s account here is not a modal one. For he does not simply combine the virtue and tracking conditions, but requires that the tracking condition holds because the virtue condition holds. Sosa’s account is explanationist, as we noted (3.2.1). 60

60. Though Sosa sometimes writes as if he required only the conjunction of both conditions: “For [in order to explain failure of knowledge in necessary truths] we can use the fact that a belief might be safe without being virtuous, while requiring beliefs to be both safe and virtuous to be knowledge” (Sosa, 2002, 275).
Nevertheless, if we leave out the requirement of an explanatory connection between the two conditions, we get an interesting double modal requirement. Let us make a few simplifying assumptions. (1) In Sosa’s (1996, 274–5) postscript, the tracking requirement “one would (most likely) believe P iff P were the case” is spelled out as Cartesian tracking, i.e. Safety and Adherence. Here we drop the Adherence requirement (as Sosa eventually did in 1999a, 376, 2000a, 40–1 and 2002, 274). (2) We assume that the virtue requirement is at least structurally equivalent to Goldman’s reliabilism.

**Safe and reliable (SAR)** \[ E_m \alpha \iff B_m \alpha \land \neg T \alpha, \]

\[ G_n \alpha \iff B_n \alpha, \]

\[ aR^1 \beta \iff p_\beta = p_\alpha, S_\beta = S_\alpha, \text{and } w_\beta \text{ is close}_1 \text{ to } w_\alpha. \]

\[ aR^2 \beta \iff w_\beta \text{ is close}_2 \text{ to } w_\alpha. \]

We leave it open whether the range of worlds relevant to virtue is different from the range relevant to safety, that is, whether closeness\(_2\) and closeness\(_1\) are identical. (One could imagine that safety depends on a much narrower set of alternatives to actuality than reliability, as with Plantinga’s (1997) distinction between the “maxi-environment” and the “mini-environment” of a belief.) Structurally, the account is almost the same as Goldman’s double reliability account. “Local” reliability or safety concerns the avoidance of error on a certain proposition; “global” reliability or virtue concerns the avoidance of error by any subject on any proposition. The only difference is that Goldman’s “local” reliability is not (or at least not clearly) subject-centred.\(^{61}\)

### 3.8 Discussion of the results

Results are summarised in tables 3.2–3.4 below. The view defended in this thesis, method infallibilism, is listed for comparison.

---

\(^{61}\) *Revision note.* I originally attributed the safety and virtue view to Pritchard (2010) as well. Duncan Pritchard’s recent formulations clarifies that he does not hold the safety and virtue view. Rather, he argues that knowledge requires a belief that is safe (and true) *because* virtuous. The view is explanationist, and thus falls without the scope of
### Table 3.2: Simple Infallibility Requirements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True belief</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$\beta = \alpha$</td>
<td>$\surd$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Infallible belief</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Goldman’s Relevant alternatives (and Lewis’s contextualism)</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $\beta$ is a relevant alternative to $\alpha$ and $\beta$ is perceptually equivalent to $\alpha$.</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, $w_\beta$ is at least as close to $w_\alpha$ as any $\neg p_\alpha$ world.</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Nozickian sensitivity (see app. ??)</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, $w_\beta$ is in the $\neg p_\alpha$ neighbourhood or closer.</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\surd$</td>
</tr>
<tr>
<td>Sosa’s Safety (variable)</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is at least as close to $w_\alpha$ as any world in which $S$ fails to believe $p_\alpha$.</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\surd$</td>
</tr>
<tr>
<td>Sosa’s Safety (fixed)</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is close to $w_\alpha$.</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Safety about $p$</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is close to $w_\alpha$.</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>Williamson’s Safety</td>
<td>$B \alpha$</td>
<td>$E \beta$</td>
<td>$\beta$ is sufficiently relevantly similar to $\alpha$</td>
<td>$\times$</td>
<td>$?\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$?\times$</td>
</tr>
<tr>
<td>Williamson’s revised Safety</td>
<td>$B \alpha$</td>
<td>$F \beta$</td>
<td>$\beta$ is sufficiently relevantly similar to $\alpha$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\surd$</td>
<td>$\times$</td>
<td>$?\times$</td>
</tr>
</tbody>
</table>

### Table 3.3: Parametric infallibility and reliability requirements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Infallibly based belief</td>
<td>$B b \alpha$</td>
<td>$B b \beta \land \neg T \beta$</td>
<td>$p_\beta = p_\alpha$</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Armstrong’s simple nomological req.</td>
<td>$B a \land C a$</td>
<td>$C \beta \land E \beta$</td>
<td>$p_\beta = p_\alpha$ and the laws of $\alpha$ hold in $\beta$</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Armstrong’s nomological req.</td>
<td>$B a \land C a$</td>
<td>$C \beta \land E \beta$</td>
<td>$\exists x \exists y (p_\beta = p_\alpha[x/y])$ and the laws of $\alpha$ hold in $\beta$</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Method infallibilism</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$\beta$ is close to $\alpha$</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Quasi-Nozickian sensitivity with methods</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$\alpha R_m \beta$ iff $p_\alpha = p_\beta$ and $w_\beta$ is at least as close to $w_\alpha$ as any world in which $p_\alpha$ is false and $S_\alpha$ uses $m$ to arrive at a belief whether $p_\alpha$</td>
<td>√</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Revised Nozick’s sensitivity with methods</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, $w_\beta$ is at least as close to $w_\alpha$ as any $\neg p_\alpha$ world.</td>
<td>√</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Basis-relative safety</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$p_\alpha = p_\beta$, $S_\alpha = S_\beta$, $w_\beta$ is close to $w_\alpha$.</td>
<td>√</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>Sosa’s reliable-indication safety</td>
<td>$B l \alpha$</td>
<td>$L \beta \land \neg T \beta$</td>
<td>$p_\alpha = p_\beta$, and $w_\beta$ is close to $w_\alpha$</td>
<td>√</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Goldman’s process reliabilism</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$w_\beta$ is close to $w_\alpha$.</td>
<td>x</td>
<td>x</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Pritchard’s most-worlds safety</td>
<td>$B m \alpha$</td>
<td>$B m \beta \land \neg T \beta$</td>
<td>$p_\beta = p_\alpha$, $S_\beta = S_\alpha$, and $w_\beta$ is close to $w_\alpha$</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldman's double reliability</td>
<td>$B\alpha$</td>
<td>$B\alpha \land \neg T\beta$</td>
<td>$p_\beta = p_\alpha$, $\beta$ is a relevant alternative to $\alpha$ and $\beta$ is perceptually equivalent to $\alpha$.</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>most:</td>
<td></td>
<td>$\beta$ is close to $\alpha$</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pritchard's double safety</td>
<td>$B\alpha$</td>
<td>$B\alpha \land \neg T\beta$</td>
<td>$p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is very close to $w_\alpha$.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>most:</td>
<td></td>
<td>$p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is close to $w_\alpha$.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety and Virtue</td>
<td>$B\alpha$</td>
<td>$B\alpha \land \neg T\beta$</td>
<td>$p_\beta = p_\alpha$, $S_\alpha = S_\beta$, and $w_\beta$ is close to $w_\alpha$.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>most:</td>
<td></td>
<td>$w_\beta$ is close to $w_\alpha$</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The tables highlight what I take to be potential sources of problems for the requirements. So a check (✓) indicates a point to pay attention to. (Unsurprisingly, I have made sure that my view does not check any box.) Whether the point is actually a problem depends on further considerations. For instance, I think that fallible requirements are unable to rule out Gettier problems. One may not see that as a problem, if one puts the requirement forward merely as a necessary condition on knowledge. Moreover, some features are only typically problematic. For instance, I think that non-parametric requirements are typically reducible to independent conjuncts and that they face a secondary type of Gettier problem. But Williamson’s safety requirement does not face this problem, even though it is non-parametric. Bearing these points in mind, let me review the features of interest.

3.8.1 Proposition- and subject-centring

Proposition- and subject-centred requirements only look at alternative cases in which the same proposition and the same subject are involved, respectively.

**proposition-centring** \( R \) is such that \( p_\beta = p_\alpha \) for any \( \alpha, \beta \) such that \( \alpha R \beta \).

**subject-centring** \( R \) is such that \( S_\beta = S_\alpha \) for any \( \alpha, \beta \) such that \( \alpha R \beta \).

The analogous notion of time-centring is interesting, but I have left it out here. Subject-centring is discussed in appendix E.

Most modal requirements in the literature are proposition-centred. Safety *about* \( p \) is quasi-centred. Goldman’s reliabilism is not proposition-centred, but it is fallibilist; and the same is true of the virtue component of the Safety and Virtue account. Armstrong’s nomological infallibilism and Williamson’s safety are the only non-centred infallibilist conditions.

Yet proposition-centring is a bad idea. Proposition-centred requirements are trivially satisfied by necessary truths. That is not a side issue but rather a symptom of the fact that such requirements are under-generalisations, as I argue in the next chapter (4.1).

the modal requirements we examine here.
3.8.2 Variability

Let me expand the characterisation of variability. Given a modal requirement involving a relation $R$ between cases, and an arbitrary case $\alpha$, we define:

**relevant worlds** The set $R^R_{\alpha}$ of relevant worlds for $\alpha$ in view of $R$ is the set of all worlds $w_\beta$ such that $\alpha R \beta$ for some $\beta$.

$$R_{\alpha} = \{w : \exists \beta(\alpha R \beta \land w = w_\beta)\}.$$  

In modal requirements, the relation $R$ selects relevant alternative cases: that is, for each case $\alpha$, it selects a range of cases in which the existence of error is relevant to whether the requirement is satisfied in $\alpha$. The set $R^R_{\alpha}$ collects the worlds in which $R$ locates relevant alternative cases for $\alpha$: a world is in $R^R_{\alpha}$ if there is some case $\beta$ in that world such that $\alpha R \beta$.

Now I call a requirement *variable* iff the sets $R^R_{\alpha}$ depend on the particular proposition involved in $\alpha$. We define:

**variability** $R$ is *variable* iff there is some pair of cases $\alpha$, $\beta$, such that $w_\alpha = w_\beta$, $S_\alpha = S_\beta$, $t_\alpha = t_\beta$, but $R^R_{\alpha} \neq R^R_{\beta}$.

(Suppose that there is such a pair. Since $R^R_{\alpha} \neq R^R_{\beta}$, $\alpha \neq \beta$. Since $w_\alpha = w_\beta$, $S_\alpha = S_\beta$, $t_\alpha = t_\beta$, $p_\alpha \neq p_\beta$: the two cases differ only by the proposition that is singled out in them.)

Proposition-centring and subject-centring on a relation $R$ do not entail variability — contrary to what one may at first think. To see this, it is important to remember that there is no constraint on the world / subject / propositions combinations that constitute a case. In particular, it is not required that the subject of the case exists or believes the proposition of the case (3.1.1). For illustration, consider $R$ such that $\alpha R \beta$ iff $p_\beta = p_\alpha$ and $S_\beta = S_\alpha$. Let $\alpha$, $\beta$ be any two cases who differ only by their target propositions, $p_\beta \neq p_\alpha$. The relevant cases for $\alpha$ and $\beta$ are distinct: for any $\gamma$, if $\alpha R \gamma$, then $\neg \beta R \gamma$. But the *worlds* of their relevant cases are the same: for any case $\gamma$ such that $\alpha R \gamma$, there is a case $\gamma^*$ which is just like $\gamma$ except that its target proposition is $p_\beta$. So $\beta R \gamma^*$. Thus $w_{\gamma^*} = w_\gamma \in R^R_{\alpha}$ and $w_{\gamma} = w_{\gamma^*} \in R^R_{\beta}$, so both relations map $\alpha$ and $\beta$ to the same set of *worlds* — in this particular case, the set of all possible worlds.
It follows that any relation \( R \) definable as:
\[
\alpha R \beta \text{ iff } p_\beta = p_\alpha, S_\beta = S_\alpha \text{ and } w_\beta \text{ is close to } w_\alpha.
\]
is not variable. Fixed safety (of the simple and basis-relative kind) are paradigmatic examples.

Requirements that characterise the set of relevant \textit{worlds} for \( \alpha \) as a function of \( p_\alpha \) are variable: a case \( \beta \) that differs from \( \alpha \) only with respect to its target proposition may in principle be mapped by \( R \) to a different set of worlds. The paradigmatic example here is Nozick’s sensitivity.

Requirements that directly rely on a relation of closeness between \textit{cases}, such as Goldman’s relevant alternatives (on my formulation), Williamson’s safety and method infallibilism are a less straightforward matter. It is in principle possible for such requirements to be variable. Let me illustrate with Goldman’s requirement. Consider a subject \( S \) who believes that \( p \) and that \( q \) (at a world \( w \) and time \( t \)). This gives us \textit{(inter alia)} two cases: \( \alpha \) with \( p_\alpha = p \) and \( \beta \) with \( p_\beta = q \). Say that \( p \) is a proposition about what \( S \) sees and \( q \) a proposition about what she hears. Is “perceptual equivalence” defined in such a way that all perceptual equivalents to \( \alpha \) are perceptual equivalents to \( \beta \), on Goldman’s view? I have assumed that that was the case. The result is close to Lewis’s view: perceptually equivalent cases are cases where the \textit{overall} experience of the subject is equivalent. But one may develop Goldman’s requirement in another direction, with the equivalents of \( \alpha \) being the \( p_\alpha \)-relevant ones (say, cases in which the subject has a relevantly similar \textit{visual} experience) and the equivalents of \( \beta \) being the \( p_\beta \)-relevant ones (say, cases in which the subject has a relevantly similar \textit{auditive} experience). Then \( \alpha \) and \( \beta \) may have alternatives in distinct sets of worlds, even though they are cases made of the same subject, time and world. Williamson’s safety and methods infallibilism could equally be developed in a variable way.

Variability is typically incompatible with the closure of knowledge under competent deduction. When the premises and conclusion of a deduction are distinct propositions (which is often so for non-boring deductions), variable requirements may assign different sets of relevant
3.8. Discussion of the results

worlds to premise and conclusion. Then the fact that a subject avoids error in the set relevant to the premise need not entail that the subject avoids error in the set relevant to the conclusion. Nozick’s sensitivity is a well-known illustration of this. Time-relative safety is another one, as we argue in appendix D.
This chapter defends the main thesis of this work:

**Method infallibilism**  
S knows that \( p \) if and only if S believes that \( p \) on the basis of a method that could only yield true beliefs.

My defence is structured as follows. I argue that a modal account of knowledge should satisfy two constraints: it should invoke methods (4.1) and it should be infallibilist (4.2). Method infallibility is thus necessary for knowledge. The Methods Constraint faces the so-called “generality problem”, and the Infallibility Constraint seems to lead to unacceptable scepticism. I defend the constraints against both objections (4.3). I then discuss three objections to the idea that method infallibility is not sufficient, and find them wanting. The thesis that method infallibility is sufficient is not refuted (4.4).

The argument is conditional and abductive. Conditional, because it assumes that a modal account of knowledge can and should be given. I have given methodological reasons for the assumption (sec. 3.2), but they are less than compelling. Abductive, because it is only confirmed by the failure of objections to it.

### 4.1 The Method constraint

The Method Constraint is the following:
Methods constraint  A satisfactory modal requirement on knowledge should require avoidance of errors based on the same method, whether or not the errors involve the same proposition or belief.

The main argument for the Methods constraint is the necessary truths problem. It is well known that some modal requirements are trivially satisfied by any belief in a necessary truth (Nozick, 1981, 186, McGinn 1984/2002, 13–4, Sosa, 2002, 275, Pritchard, 2007, 280). ¹ But the significance of that problem has been underestimated (though see McGinn 1984/2002, 15–7). When fully appreciated, the problem leads to the replacement of belief-based requirements with method-based ones.

4.1.1 The necessary truths problem

Let us take a simple safety requirement on knowledge as our starting point:

Simple safety  S knows that p only if S could not easily wrongly believe that p.

We say that S could easily wrongly believe that p iff there a suitably close possibility or case in which S believes p and p is false. Simple safety is a proposition-centred requirement: it obtains in a case α depending on whether there are errors that p in relevant alternative cases (sec. 3.8.1).

Consider the following pair of cases.

Prime Number  Primo has a mistaken way of evaluating whether a number higher than 19 is prime: he adds up its digits and declares the number prime iff the sum of its digits is prime. This gives false answers for many numbers, but true ones for some. I randomly pick up the number 47, write it on a piece of paper and show it to

---

¹ More generally, they are either trivially satisfied, or impossible to satisfy, or inapplicable. The safety requirement (Bp□ → p) is trivially satisfied whenever p is a necessary truth. The sensitivity requirement is either trivially satisfied, or inapplicable if one thinks that subjunctive conditionals with impossible antecedents have no truth value. Variants of the sensitivity conditional are trivially impossible to satisfy: ¬(¬p□ → Bp), it is not the case that if p was false S would believe p (Williamson, 2000, 149–80).
4.1. The Method constraint

Primo.

Rigid Variant ($\alpha$): I ask Primo whether ($p_\alpha$) that number is prime.

Non-rigid Variant ($\beta$): I ask Primo whether ($p_\beta$) I have written a prime number.

In both cases Primo answers (rightly) “yes”.² Safety is satisfied in the rigid case but not in the non-rigid one. $p_\alpha$ is the proposition that 47 is a prime number. There is no possible case in which 47 is not prime. Thus there is no possible case in which Primo believes that 47 is prime while it is not. So simple safety holds in $\alpha$, no matter how closeness is fixed. By contrast, given the setup, I could easily have written another number on the paper, say 49. If I had shown that number to Primo, he would have mistakenly judged it prime. So there is a close case in which Primo believes that $p_\beta$ but $p_\beta$ is false. So simple safety does not hold in $\beta$.

Primo clearly does not know that the number is prime, even though he has a true belief that it is. Simple safety explains why he fails to know in $\beta$. The truth of Primo’s belief is too accidental in that case, and the safety requirement brings the accidentality out. But it fails to explain why he fails to know in $\alpha$.

Do the cases pose a problem? Simple safety was only put forward as a necessary requirement on knowledge. Primo’s case $\beta$ shows that it is not sufficient, but it was not never claimed to be so. That is true, but it does not get simple safety off the hook, for two reasons. First, simple safety creates an unexpected asymmetry between the two cases (McGinn 1984/2002, 15). The difference between the two propositions appears irrelevant to Primo’s knowledge of them. His epistemic position with respect to one is just the same as his epistemic position with respect to the other. Yet simple safety entails that the first case satisfies a requirement on knowledge that the second does not. Second, since Primo fails to know in the first case, there must be some other requirement $C$ on knowledge

². The labels come from the fact that $\alpha$, but not $\beta$, involves rigid designation of a number. See McGinn (1984/2002, 14) for a similar case. (I add the rigid/non-rigid variants and remove the deduction step.)
that he fails to satisfy. But if $C$ explains why Primo fails to know in $\alpha$, it is to be expected that $C$ also explains why Primo fails to know in $\beta$, given the similarity between the cases. So simple safety will either be an under-generalisation of $C$ or made redundant by it.

(Two caveats to the argument. First, it may be that the only natural requirement $C$ we can find to explain $\alpha$ is that of knowing itself. Simple safety would remain the best non-tautologous requirement we can put forward. To this I will reply by showing that we can do better. Second, a requirement $C$ may replace simple safety without making it explanatorily redundant. $C$ may be, for instance, the requirement that a case be \textit{not relevantly like a case of non-safety}.\textsuperscript{3} Since $\alpha$ is relevantly like $\beta$, and $\beta$ is a case of non-safety, both $\alpha$ and $\beta$ fail to satisfy this requirement. But since the requirement is formulated in terms of simple safety, the latter still has an explanatory role to play. In reply, I cannot \textit{a priori} rule out the possibility of such a requirement. But I doubt that one can be put forward which is both satisfactory and which cannot be replaced by a requirement that does \textit{not} appeal to simple safety. For instance, the one just envisaged has no advantage over the simpler requirement that a case be \textit{not relevantly like a case of error}.)\textsuperscript{4}

Are necessary truths a special case? One lesson of Kripke (1980) is that necessary truths are not confined to special domains like mathematics, metaphysics or fundamental sciences. Accordingly, the problem raised by

\textsuperscript{3} Thanks to John Hawthorne for the suggestion.

\textsuperscript{4} Revision note. Igor Douven made a further objection to the argument. While the argument indicates that in cases involving proper names, rigid designators, or natural kinds predicates, simple safety will be made redundant by a condition that covers both cases in the symmetric pair, it is still possible that the condition fails to deal with other cases that simple safety would account for.

Now, a condition that would cover both cases of the pair (and of others like it, see below) does not merely cover cases involving a rigid designator: for it also covers the non-rigid variant of the case. So if anything, it may be that the condition will only apply to cases for which a rigid variant exist. But using an actuality operator, we show below that for any case, a rigid variant exists. So any condition that meets the challenge of such pairs should cover any cases that simple safety covers. Note that, since a condition cannot covers all the cases covered by simple safety without entailing simple safety, that means that the putative condition will entail simple safety. The method requirement we put forward is a case in point.
4.1. The Method constraint

the Prime Number case generalises (McGinn 1984/2002, 14–5). Consider first a pair of cases based on Ginet-Goldman’s fake barns case (Goldman, 1976, 772–3): 5

FAKE BARNs  Bernard drives by the countryside with his son, pointing at various objects in the field. He sees a barn, points it to his son, and says:

Rigid Variant $\alpha$: “That building is a barn!” ($p_\alpha$)

Non-rigid Variant $\beta$: “The building I am pointing at is a barn!” ($p_\beta$)

The building is a barn. However, fields in the area are full of papier-mâché barn facades that Bernard would have mistaken for barns.

Let $b$ be the building Bernard is pointing at. $p_\alpha$ is the proposition that $b$ is a barn. Arguably, there is no possibility in which $b$ is not a barn. If a barn facade made out of papier-mâché had been built at the time and place $b$ was built, it would not have been $b$, but another building. So there is no possibility in which $p_\alpha$ is false; hence no possibility in which Bernard believes that that building is a barn while it is not. 6 By contrast, if details are filled in properly, there is a close possibility in which Bernard would be pointing to a barn facade, and in which he would nevertheless believe that it is a barn. So here as well, simple safety holds for $p_\alpha$ and not for $p_\beta$.

But again, the variants appear perfectly symmetric.

The argument relies on two assumptions. First, I am assuming a measure of essentialism such that a barn could not have been a barn facade. For instance, given essentiality of origins, a building made out of different materials than $b$ would not be identical to it. But whatever one thinks about the essence of barns, analogous cases can be formulated for anybody who accepts some measure of essentialism. 7 Thus if one thinks that animals belong to their species essentially, we can build an analogous

---

5. The case was suggested to Goldman by Carl Ginet.
6. This is a bit quick. Perhaps there are possibilities in which $b$ still exists and is not a barn. For instance, $b$ may later be converted into a house, etc. But we may assume that any of these possibilities is either distant or such that Bernard would not then mistakenly believe that $b$ is a barn.
7. Except perhaps somebody who thinks that things only have their appearance essentially.
case involving fake sheep. Second, I assume that the demonstrative “that building” is a rigid designator (Kaplan, 1989; Récanati, 1993). Again, if that is contested, the argument can be recast as long as one accepts that some expressions rigidly designate ordinary objects. For instance, the case can be recast with proper names. To illustrate both points:

**Fake Sheep** Bertha sees a sheep, baptises it “Ajax”, points it to her son and says:

*Rigid Variant* $\alpha$: “Ajax is a sheep!” ($p_\alpha$)
*Non-rigid Variant* $\beta$: “The animal I am pointing at is a sheep!” ($p_\beta$)

Ajax is a sheep. But unknown to Bertha, many other animals in the field are Greek soldiers disguised as sheep in fear of a Cyclops.

Thus it cannot be said that simple safety solves the fake barn case. It only solves variants of the case involving non-essential properties and non-rigid designators.  

The problem is further generalised to cases involving predicates. First, there is the predicate analogue of rigidly referring individual terms:

**Fool’s Gold** A small tribe discovers a metal unknown to them in the ground, and baptises it “gold”. However, the area also contains much of an indistinguishable (to them) substance that is not a metal — call it fool’s gold — which they might have just as well hit upon.

*Rigid variant*: they believe that gold is a metal.

*Non-rigid variant*: they believe that the substance they found is a metal.

Again, there is a close possibility in which the second belief is false, but none in which the first is. Second, there are analogous cases involving

---

8. Perhaps fake-barn style of cases involving rigid designation of places, times or abstract entities can also be constructed. I leave such constructions to others.

9. See e.g. Pritchard (2005, 162): “We can explain this intuition in terms of our account of veritic epistemic luck by noting that there are going to be a great many nearby possible worlds where Henry forms the same belief on the same basis (by simply looking at the ‘barns’) and yet his belief is false.” Whether Henry/Bernard forms the same belief at nearby worlds depends on whether beliefs have their content essentially. If they do, Bernard forms the same belief only in the non-rigid variants. See Weatherson (2004) for a notion of “belief safety” that allows beliefs to have different contents at close worlds. However, as I argue below, this narrow content approach is insufficiently general.
vague predicates:

**Small number** Let Sam be in a conversational context in which “small”
takes a semantic value such that 11 is the greatest natural number
to satisfy “small” in that context (if epistemicism about vagueness
is true), or such that 11 is the greatest natural number to satisfy all
acceptable precisifications of “small” (if supervaluationism is true).
Sam believes that 11 is small.\(^\text{10}\)

Let \(S\) be the property (if epistemicism is true), or the set of properties
(if supervaluationism is true), that “small” expresses in Sam’s context.
On the epistemicist story, the proposition that Sam believes is \(S(11)\), and
it is a necessary truth. On the supervaluationist story, Sam’s beliefs is
associated with a set of propositions, each of which is a necessary truth.
On any of the two accounts, there is no close possibility in which Sam’s
belief is associated with the same propositions and false or not supertrue.
If Sam has no particular insight into the cutoff points, there are close
possibilities in which “small” has a slightly different semantic value and
she acquires a false or not supertrue belief. But that possibility is not
relevant to simple safety.

The problem can be dramatically generalised if we help ourselves with
a slightly artificial device. Let \(\text{actually}\) be a rigid designator for the actual
world.\(^\text{11}\) \(\text{Actually } p\) uttered in a world \(w\) is true at any world \(w’\) iff \(p\) is
ture in \(w\). For any true proposition \(p\), \(\text{actually } p\) is a necessary truth. Now
take any case of true belief that \(p\) (for some proposition \(p\)) that fails to
satisfy simple safety. There is a case just like it except that the proposition
believed is \(\text{actually } p\). The latter one satisfies simple safety, even thought
the subject is in the same epistemic position with respect to \(\text{actually } p\) as
she is with respect to \(p\) in the original case.

The problem is not limited to strict necessities either. Like other
safety conditions, Simple safety relies on some notion of easy possibility
\(^{10}\) See Fine \((1975)\) for the supervaluationist account and Williamson \((1994)\) for the
epistemicist one.

\(^{11}\) I doubt that the English adverb “actually” is such a designator. See Stephanou
\((2010)\) for a recent discussion.
(Sainsbury, 1997). Easy possibility is stronger than strict possibility: if it is easily possible that \( p \), then it is strictly possible that \( p \), but the converse is not valid. Call \emph{weak necessity} the dual of easy possibility: it is weakly necessary that \( p \) iff it is not easily possible that \( \neg p \). Weak necessity is weaker than strict necessity: if it is strictly necessary that \( p \), it is weakly necessary that \( p \), but the converse is not valid. In possible worlds terms, the notions can be represented by introducing a relation of closeness between worlds. We say that at a world \( w \) it is easily possible that \( p \) iff there is a world \( w' \) close to \( w \) in which \( p \) occurs; and that is is weakly necessary that \( p \) at a world \( w \) iff in any world \( w' \) close to \( w \), \( p \) occurs. Weakly necessarily true propositions raise the same problem for simple safety.\(^{12}\)

For illustration, suppose that it is a weak necessity that Ringo wears a wedding ring if he is married. It is not a strict necessity: if Ringo’s upbringing had been different, say, he might have been liable not to wear a ring while married. But as things are, he would not fail to wear a ring if he was married.

\textbf{Wedding ring}  Ringo would wear a ring if married, but he is not. However, many of his married colleagues do not wear wedding rings. A foreign client visits Ringo’s office while Ringo is at the reception desk and notices that he wears no ring.

\textit{Rigid variant} \( \alpha \): The client forms the belief that Ringo is not married.

\textit{Non-rigid variant} \( \beta \): The client forms the belief that the reception clerk is not married.

We may fill in the case so that there is a close possibility in which one of Ringo’s ringless married colleagues is at the reception desk, and the client wrongly believes that they are not married. Again, simple safety is not satisfied in \( \beta \) but satisfied in \( \alpha \), though the epistemic position of the client is the same towards both propositions.

There are a few drastic moves one can make to deny that simple safety is satisfied in our cases. One is to argue that there \emph{is} a possibility in which

\(^{12}\) See sec. 3.7.4 and appendix A.
47 is not prime, and analogously for the other cases. “Impossible worlds” could be introduced in which contradictions are true, or bivalence fails, or some metaphysical or nomological necessities do not hold. These are obviously not possibilities in the metaphysical or nomic sense, but they could be possibilities in some epistemic sense. Assuming some such notion can be devised, the question arises: of which subject are these the epistemic possibilities of? There are broadly two options. First, they are the epistemic possibilities of the subject to whom knowledge is ascribed. Simple safety becomes roughly: $S$ knows that $p$ iff it is not (easily) epistemically possible for $S$ that $S$ believes that $p$ and $p$ is false. This deprives safety of most of its explanatory power. Many take epistemic possibility to be defined as the dual of knowledge.\textsuperscript{13} Thus saying that Bernard’s belief that $p_a$ is unsafe would just be a fancy way of saying that he does not know that $p_a$. Others take epistemic possibility to be compatibility with one’s evidence, on some internalist notion of evidence that is not equated with knowledge.\textsuperscript{14} It then becomes unclear whether there is an easy epistemic possibility for Bernard that the building he is pointing at is not a barn. Second, the epistemic possibilities could be the ones of whoever ascribes knowledge. But then simple safety is again satisfied in our examples, since we know that the building Bernard is looking at is a barn and that 47 is prime. There is, additionally, the general worry that allowing impossible worlds to be relevant will have unwanted sceptical consequences, since about anything can happen in them.

Another drastic move is to reformulate safety in terms of narrow content, along the following lines:

**Narrow content safety** If $S$ believes that $p$ with narrow content $n$, then $S$’s belief is safe iff $S$ could not easily have had a false belief with narrow content $n$.

\textsuperscript{13} That is, $p$ is epistemically possible to $S$ iff $S$ does not know that not-$p$; or iff $S$ is not in position to know that not-$p$. See e.g. Stanley (2005a).

\textsuperscript{14} See e.g. Fantl and McGrath (2009, 11–2).
Methods infallibilism

(Analogous proposals can be made with belief vehicles or with narrow notions of modes of presentation.\(^{15}\)) If an adequate notion of narrow content can be devised, it will probably dispose of some of our cases. But I will not be sufficiently general to deal with all of them. In the prime number case, Primo’s belief that 47 is prime will count as unsafe only if it is argued that, say, his belief that 49 is prime would have the same narrow content. I doubt that any reasonable notion of narrow content will warrant that.

Leaving drastic moves aside, the cases show that simple safety treats epistemically symmetric cases in an asymmetric way. Simple safety is unsatisfactory, because it is either redundant or an under-generalisation.

### 4.1.2 Method safety

The necessary truths problem affects a wide range of modal requirements. The argument does not depend on how closeness is determined. It can be extended to the infallible belief, Goldman’s relevant alternatives and sensitivity requirements, which only differ from simple safety by the way they select close cases.\(^{16}\) Relativisation to bases does not help: if there is no close possibility in which \(S\) believes wrongly that \(p\), there is no close possibility in which \(S\) wrongly believes that \(p\) on the same basis. Nozick (1981, 186–7) tried to use his adherence requirement to avoid the problem: one fails to know a necessary truth \(p\) if one could easily fail to believe \(p\) rightly while using the same method to tell whether \(p\).\(^{17}\) But that does not avoid the problem either. In our prime number case, Primo’s method is robust: Primo would form the belief that 47 is prime if asked whether it is. In our fake barn case, Bernard would form the belief that \(b\) is a barn, if he were to judge on the basis of its looks as he did. Switching to safety about \(p\) does not help either, for the same reason.\(^{18}\)

The problem affects any requirement for which errors are relevant only

---

16. These requirements are defined in 3.4.
17. Adherence is defined in 3.5.3.
18. Safety about \(p\) is defined in sec. 3.4.5.
if they involve \textit{the same belief} — whether that means a belief with the same broad content or the same narrowly-individuated belief state. In particular, it affects all proposition-centred requirements (McGinn 1984/2002, 16–7). Most modal requirements discussed in chapter 3 face it. The only ones that do not are Williamson’s safety, Armstrong’s nomological account, Goldman’s process reliabilism and double reliability, and the Safety and Virtue account.

The problem has been properly diagnosed by McGinn (1984/2002, 16–8), in my view. Primo fails to know that 47 is prime because his belief is based on a wrong method. The method is wrong because it leads to false beliefs in many cases. For instance, it would lead him to the false belief that 49 is prime. Both his beliefs in the rigid and non-rigid variants are based on that method, and both fail to be knowledge for that reason. Say that a method is \textit{unsafe} if it could easily lead to false beliefs — on some suitable notion of easy possibility. The suggestion is:

\textbf{Method safety} \ S \text{ knows that } p \text{ only if } S \text{'s belief that } p \text{ is based on a method that could not easily lead to false beliefs.}

What explains that Primo fails to know is not an \textit{unsafe belief}, but an \textit{unsafe method}.  

\[19. \text{ If some adherence or avoidance of ignorance requirement is necessary for knowledge as well, an analogous problem will affect any requirements for which relevant ignorance possibilities only involve the absence of the same belief.}\]

\[20. \text{ See tables 3.2-3.4 pp. 147-149.}\]

\[21. \text{ McGinn (1984/2002, 15) mixes what I take to be the right diagnosis with a wrong one, according to which the necessary truth problem arises from the } \textit{counterfactual} \text{ character of the analysis:}

\text{[...] the same kind of thing is going wrong in cases in which (e.g.) a true existential belief is inferred from a false singular belief (in a justified way), whether the former belief is in a necessary or a contingent truth; so we do not want our account of what is going wrong to treat these cases differently. If this uniformity constraint is accepted, it shows that the failure of the tracking analysis to deal with knowledge of necessary truths reflects, and is symptomatic of, a deeper and more general inadequacy: that counterfactual dependence is not the right way to handle Gettier cases and so does not give the correct analysis of knowledge.}\n
See also (1984/2002, 18) the (surprising) claim that proposition-\textit{uncentred} ("global") tracking cannot deal with the necessary truth problem. McGinn consequently avoids modal notions altogether and defends an account in terms of actualist propensities. The neces-
But what do we mean by “method” here? In my use, as in Nozick’s (1981, 184), “a person can use a method without proceeding methodically, and without knowledge or awareness of the method he is using.” But as opposed to Nozick and McGinn (1984/2002, 17–8), I do not assume that we have a prior grasp on the notion. Rather, we infer from Primo’s case that his possible error about 49 and his true belief about 47 are equivalent or similar enough in epistemically relevant respects, even though they do not involve the same belief nor the same proposition. The respects have to do with the subject’s mental states, cognitive processes, but the may also have to do with the broader situation she is in, e.g. her environment and how she is related to it. Call this the cognitive set up in which a belief is situated. Two cases share a method if they have sufficiently similar cognitive set ups, that is, if there is an epistemically relevant class of cases to which both belong.

Most existing notions of safety already include a reference to some notion of method or basis of belief (Williamson, 2000, 128; Sosa 1999a, 378; 1999b, 149; 2005, 156; Pritchard, 2005, 152–5). We will get a better grasp on the work the notion of method is meant to do by comparing method safety with the only account that does not make reference to methods or basis of belief, Williamson’s Similar Cases Safety.

### 4.1.3 Relevantly similar cases safety

Williamson’s safety (2000, 100, 123, 147) is the only modal requirement to avoid the notion of method altogether:

**Williamson’s safety** S knows that p only if S avoids false belief in sufficiently similar cases.

A counterfactual such as “If S were to form another belief on the same basis, that belief would be true”, while not unproblematic, does not obviously face the necessary truth problem. More generally, methods-based modal requirements can avoid it.

22. Thus a method in my use of the term need not be “a set of rules such that one is, at least in principle, always in a position to know whether one is complying with them” (Williamson, 2008, 277). Williamson’s notion of method is rather intended to capture an internalist conception of what it is to proceed methodically.
Plausibly, error need only be avoided in relevantly similar cases, that is, cases that are similar in respects relevant for knowledge. In case $\alpha$, I see that there is an apple on the table, and the conditions are as good as may be. Case $\beta$ is extremely similar to $\alpha$: every molecule in the world is positioned as it is in $\alpha$, except for a few changes in my brain so that I believe instead that there is no apple on the table. Is $\beta$ sufficiently similar to $\alpha$? If yes, I do not know in $\alpha$; but that seems implausible. (Note in particular that $\beta$ need not be a metaphysically “close” possibility in any sense — for instance, it need not have had the slightest chance to occur in $\alpha$. Whatever closeness between cases there is here only depends on similarity considerations.) If no, only cases that are even more similar to $\alpha$ than $\beta$ is are relevant to whether I know. That would seem to imply, for instance, that what would go on if there was a pear on the table instead of the apple does not matter; but it certainly does. The right reply is, of course, that $\beta$ is not sufficiently relevantly similar to $\alpha$. For purposes of knowledge evaluation, the quasi-perfect match between $\beta$ and $\alpha$ in the distribution of physical matter in the universe is significantly outweighed by the fact that some neural quirk produces an opposite belief in $\beta$.

The notion of relevance is flexible enough to ensure that Williamson’s safety is (at last to a large extent) extensionally adequate. The flexibility makes it less informative. Still it says enough to ground Hintikka-style models of knowledge in which the accessibility relation ranges over cases. With the additional claim that the relevant similarity relation is non-transitive, that is sufficient for Williamson’s anti-luminosity argument (Williamson, 2000, chap. 5). That being said, it is not as if we had no insight into what the relevant respects of similarity are. (The idea that the similarity relation is not transitive, for instance, derives from such insights.) The insights are based on our intuitive judgements about

---

23. See e.g. Williamson (2000, 100): “we need not even assume that we can specify the relevant degree and kind of similarity without using the concept knows.”

24. In light of the previous discussion (4.1.1), extensional adequacy does not only require that all instances of knowledge satisfy the requirement, but also that the requirement does not create unwarranted asymmetries between cases of non-knowledge that satisfy it and cases that do not.

25. See Williamson (2000, 100–1, 2009a, 23).
knowledge in various cases, but that does not make them unexplanatory (Williamson, 2000, 100-1).

Now Williamson writes at times as if the relevant similarity could be factorised into relevant similarity of components of cases:

It at time $t$ on a basis $b$ one knows $p$, and at time $t'$ close enough to $t$ on a basis $b'$ close enough to $b$ one believes a proposition $p'$ close enough to $p$, then $p'$ should be true. (Williamson, 2000, 102; see also Williamson, 2009c, 325).

This suggests that a case $\beta$ is close enough to $\alpha$ iff its target proposition is close enough, and its basis is close enough, and so on. But that cannot be right, for the different dimensions of closeness have to be coordinated. Consider the four following cases:

- $\alpha$: Seeing a tree from a normal distance, $S$ judges that it is $10 \pm 2$ m. high.
- $\beta$: Seeing a tree from a short distance, $S$ judges that it is $10 \pm 1$ m. high.
- $\gamma$: Seeing a tree from a long distance, $S$ judges that it is $10 \pm 3$ m. high.
- $\delta$: Seeing a tree from a long distance, $S$ judges that it is $10 \pm 1$ m. high.

If dimensions of closeness are independent, $\beta$ and $\gamma$ cannot be close to $\alpha$ without $\delta$ being close to $\alpha$. Suppose $\beta$ and $\gamma$ are close to $\alpha$. Since $\beta$ is close to $\alpha$, the proposition in $\beta$ is sufficiently close to that in $\alpha$. Since $\gamma$ is close to $\alpha$, the basis in $\gamma$ is sufficiently close to the basis in $\alpha$. Since the proposition in $\delta$ the same as in $\beta$, it is sufficiently close to that of $\alpha$; and since the basis in $\delta$ is the same as in $\gamma$, it is sufficiently close to that of $\gamma$. Hence $\delta$ is also close to $\alpha$. But depending on the details of case $\alpha$, we may very well want to count $\beta$ and $\gamma$, but not $\delta$, as close to $\alpha$. In both $\beta$ and $\gamma$, the inexactness of the proposition believed is proportioned to the distance of the tree is as it is in $\alpha$. Case $\delta$ involves believing a more exact proposition than in $\alpha$ on a weaker basis. $S$ in $\alpha$ may not be at all disposed to such rashness; and if not, $\delta$ should not be counted as close to $\alpha$. Dimensions of similarity should consequently be coordinated: whether the proposition in $\delta$ is sufficiently close to that of $\alpha$ depends on how the basis in $\delta$ differs from that of $\alpha$. 
What we need is a notion that covers the coordination within the different respects of similarity. This is what the notion of method does. Various ways to proportion one’s beliefs to one’s situation and/or stimuli make for different methods. The point above can thus be simply reformulated as follows: in \( \alpha \), \( S \) uses a method that is the same, or similar enough to, the method used in \( \beta \) and \( \gamma \) without being the same as, or similar enough to, the one used in \( \delta \). Similarity or identity in methods encompasses and coordinates similarity in proposition believed, in level of confidence, in bases (if these are distinct from methods themselves), and perhaps in other respects. I leave it open at this stage whether the dimensions of subject, time, place and world should also be integrated in methods.

### 4.1.4 Conclusion

Any satisfactory modal requirement on knowledge should take into account epistemically relevant classes of cases that may involve different beliefs or different propositions. We call such classes *methods*. Call this the Method Constraint.

Two remarks on my implementation of the Method Constraint are in order.

First, there are two ways in which to implement a method-based requirement. \(^{26}\) One is to use method similarity, the other is to use method identity:

**Method-similarity requirement** \( S \) knows that \( p \) only if \( S \)’s belief is based on a method such that all cases involving a similar enough method are cases of true belief.

**Method-identity requirement** \( S \) knows that \( p \) only if \( S \)’s belief is based on a method such that all cases involving the same method are cases of true belief. \(^{27}\)

---

\(^{26}\) Thanks to Timothy Williamson (p.c., in reply to a presentation based on my 2007) and to John Hawthorne for pointing this out to me.

\(^{27}\) “All” is replaced by “most” on fallibilist view; “only if” can be replaced by “if and
There are plausible considerations in favour of a method similarity approach. Suppose that $\alpha$ is such that all cases that are sufficiently similar in epistemically relevant respects are cases of true belief. We may say that the range of cases similar enough to $\alpha$ constitutes a method. But now we may have a pair of cases $\beta$, $\gamma$ such that: $\beta$ is just similar enough to $\alpha$ to be in the range, and $\gamma$ is only slightly different from $\beta$, but out of the range. Now $\beta$ involves a method that $\gamma$ does not involve: namely, that of being sufficiently similar to $\alpha$. Moreover, the method is infallible, by assumption. However, for all we have said, $\gamma$ may be an error case. And we may be reluctant that say that $\beta$ is a case of knowledge, given that there is a case that is very similar to it in which one is in error (see Williamson, 2000, 98–102 for such considerations and an example). But on the method identity view, we would have to say that $\beta$ is a case of knowledge. To deal with such cases, one could deny that $\alpha$ and $\beta$ share the same method, but adopt instead much finer-grained methods (at one extreme, one per case), and use a method similarity requirement.

The difference between method similarity and method identity has implications for the luminosity of knowledge (see Williamson, 2000, ch. 5 for the relevant considerations). But as long as we do not get into such issues — and we do not in the present work —, method identity seems to be at least a reasonable approximation. In the method-based models for knowledge developed in ch. 5, we assume method identity, which allows for simpler models. I will do the same in the rest of this thesis. I leave the investigation of method similarity approaches for further work.

Second, we do not require methods to be the only restriction on relevant alternative cases. Our method-based requirements have the following form, for some relevant notion of possibility:

**Method-based requirement** S knows that $p$ (if and) only if S’s belief is based on a method such that all/most possible cases involving the same method are cases of true belief.

*only if*” if one thinks method infallibility or reliability is also sufficient. The domain of possible cases we quantify over may be restricted by some notion of possibility. The latter restriction is discussed in the second point below.
4.2. The Infallibility constraint

The precision is made idle if the relevant notion of possibility is unrestricted. If any metaphysically possible case is counted as possible, the relevance or irrelevance of possible errors is entirely to be accounted for by similarity or dissimilarity in method. Williamson’s similar-case safety (see sec. 4.1.3 above and sec. 3.4.6) is an instance of this approach, albeit with a similar-method formulation instead of a same-method one. For instance, our knowledge of the future depends on the fact that the world is not suddenly going to stop going as it goes. On this approach, the fact would have to be part of some methods on which our future beliefs are based.

Alternatively, we may want to factorise the requirements into two aspects. To be relevant to whether one knows in $\alpha$, a possible error should (a) involve the same method, and (b) be a possible alternative of $\alpha$, for some restricted but natural notion of possibility. For instance, we may want to say that only cases that occur in a limited space-time region around $\alpha$ are relevant, or that only nomologically possible cases are relevant. This need not be done by “packing” the space-time location or the laws in the method used, but instead by restricting the range of possible cases in which an error on the basis of the same method is relevant. This is the approach we ultimately adopt here, though we will not present a systematic case for it. It seems to me a natural way to describe many cases (see the cases discussed below sec. 4.3.4 and the case of induction in sec. 5.3.4). Some support for it can be found in the links between the semantics of “know” and the notion of real possibility at play in the semantics of counterfactuals and modals (ch. 7).

The use of a notion of method leads to the Generality problem. We address it in sec. 4.3.

4.2 The Infallibility constraint

The Method Constraint tells us that knowledge requires avoiding error on the basis of some method, over a natural range of possible cases. The Infallibility Constraint is that there must be a method and a range such
that error is impossible on that basis in that range:

**Infallibility Constraint** Knowledge requires a substantially infallible method, that is, a method infallible over a natural range of possible cases.

A requirement is infallibilist iff it entails truth. Truth is the minimal infallibilist requirement. But truth is not sufficient for knowledge. Errors should be avoided at some other cases beyond the target one. The Infallibility Constraint says that error should be entirely avoided in some natural range of cases around the target. It is not sufficient that error be avoided in most cases within any natural range.  

The main argument for the Infallibility Constraint is that infallibility is needed to avoid both the Gettier problem and the lottery problem. Any reliability requirement presented as sufficient for knowledge in conjunction with truth is lead to ascribe knowledge in lottery cases and Gettier cases. The argument has already been put forward in the literature (Sosa, 1985, 239–40, Sturgeon, 1993, 160–1, Zagzebski, 1994, 69, Merricks, 1995, 841–5). (See appendix B, on Merrick's “Warrant entails truth” formulation and why it is unsatisfactory.)

### 4.2.1 The fallibilist dilemma

Consider the two following series of cases, due to Unger (1968, 161-2):  

**Series of card draws** A deck of cards consists of ninety-nine white cards and a black card. We shuffle them fairly and ask S to tell the colour of the card that has come on top. S knows what the deck contains and that it is shuffled fairly, but cannot see the colour of the card. Based on likelihood alone, he answers that it is a white card.

---

28. If one does not want to allow method requirements based on some restricted notion of possibility (see sec. 4.1.4), the argument of the present section is that one knows only if one’s belief is based on a method that is absolutely infallible, that is, that excludes error in any metaphysically possible cases.

29. See also Harman’s (1968) analogous contrast between a lottery case and an ordinary testimony (newspaper report) case.
We repeat the process one hundred times, and he answers likewise each time. He is wrong in one case.\footnote{The probability of the black card turning up at least once in such a series is $\approx 0.63$; the probability of it turning at least twice up is $\approx 0.26$.}

**Series of additions** S is given a hundred simple addition problems. The problems involve three small numbers, between 10 and 100. S solves them using normal paper-and-pencil methods for adding, and each time checks the results by an independent arithmetic method. He is wrong in one case.

Let me add a useful third series:

**Series of photographs** We have a series of a hundred black and white photographs of groups of about twenty adult people, in the style of school class or worker team photographs. Their quality is variable: some are much grainier than others. S is shown them and asked each time whether a certain good friend of his is in the photograph. He is wrong in one case. (In that case, he mistakenly thinks that the friend is in the photograph.)

The first series is a variant on lottery cases.\footnote{See Hawthorne (2004, 1–2) for a presentation of lottery cases and references.} The common judgement is that S does not know, in any case of the series, that the top card is white. (Whether or not he indeed believes that it is.) The second and third series are ordinary cases of mathematical and perceptual knowledge. I would not say that in *every* case of these series where S gets the answer right, he knows. But it is hard to deny that he does so in *some* of the cases of the series, provided nothing unusual is going on. At the 10\textsuperscript{th} addition problem, say, S has warmed up a bit, is still attentive enough, calculates rightly that $84 + 37 + 31$ is 152 and has checked the result. In the 10\textsuperscript{th} photo, which is in a reasonably good condition, S immediately notices the familiar face of his friend in the first row, and further examination only increases his confidence. These are paradigmatic cases of what we ordinarily call knowledge.

The contrast between the lottery case and the two others creates a dilemma (Cohen, 1988, 92; Greco, 2003, 112-3). If we require infallibility
for knowledge, we have to deny knowledge in the ordinary series. But if we say that reliability is sufficient, we have to grant knowledge in the lottery series. We can sharpen the dilemma using the apparatus of chapter 3. Prima facie, all cases in one series form a natural range, and they do not appear to involve different methods. A method infallibility requirement requires avoidance of error in all cases in the range; a method reliability requirement only requires avoidance of error in most cases in the range. The reliability requirement (in conjunction with belief and truth) cannot be sufficient, otherwise S would know in the lottery series. The infallibility requirement cannot be necessary, otherwise S would fail to know in the ordinary series. But there seems to be no modal requirement stronger than reliability and weaker than infallibility.

4.2.2 The structure in ordinary cases

What distinguishes the lottery series from the ordinary ones? An important point is that the ordinary series have a structure that the lottery one lacks (Unger, 1968, 162). Some of the photographs are more grainy than others, the faces are more recognisable in some than in others, and so on. Some addition problems are more difficult or confusing than others, and S was more attentive and less tired while doing some than others. By contrast, the lottery series is uniform with respect to epistemically relevant properties. The black card was further from the top in some cases than in others, S was less bored in some cases than in others, but that did not make him any closer to knowing that the top one was white. More generally: structured series contain natural ranges of cases in which errors are scarce and natural ranges of cases in which error is more frequent. Call the first favourable areas, and the second unfavourable areas. Of course we could single out an error-free set of cases in the lottery series. But that would be an arbitrary selection, not an epistemically natural range. Any non-arbitrary selection in the lottery series will include errors just as well.

32. Unger does not say what he means by a “more structured” case (Unger, 1968, 162), so I cannot tell if he had in mind the idea I develop here.
4.2. The Infallibility constraint

as any other; that is what makes the lottery random. By contrast, in the ordinary series, favourable areas are distinguished. We can picture the idea by imagining the series of cases as disposed in a line. In the lottery cases, errors are distributed randomly along the lines, while in the structured cases, they tend to be in some segments of it. (See also the illustrations on pp. 179–181.)

4.2.3 The Gettier problem for fallibilism

The introduction of structure opens the way for requirements that are both stronger than overall reliability and weaker than overall infallibility:

**Favourable area requirement**  S knows that \( p \) in a case only if S’s method is reliable over the series and the case is in a favourable enough area.

How favourable is favourable enough? Is it required that S’s method be infallible in the favourable area? Or is some high measure of reliability sufficient? The later proposal runs into two problems.

First, we can build lottery series with reliably favourable areas. Consider a variant of our lottery series in which the size of the deck varies, but always contains a unique black card. S is aware of the changes of size and answers as before. Errors tend to be grouped in cases in which the deck is small; when the deck is big, S is almost always right. If sufficient for knowledge (in conjunction with truth), the favourable requirement would entail that S knows in cases in which the deck is very big. But S does not know that the top card is white, even when the deck is very big. One way to press this judgement is to focus on the favourable area in which the deck is, say, 1000 cards. We may presume that there are a few cases in the area where S is wrong.\(^{33}\) The situation in that area is the same as in the original series, namely, we have an *unstructured* area in which errors are randomly distributed.\(^{34}\)

---

33. I am assuming a suitably long series. In cases in which one plays only a few times, the relevant “series” of cases is spread over counterfactual possibilities.

34. “Randomly” is loosely used here. What is meant is that errors are not grouped in an epistemically natural sub-range of cases.
(What if the deck is extremely big, say a billion cards? Intuitions are less clear, and I think that these sort of cases raise deep issues that I will treat separately: see sec. 4.3.5. Here I can simply point out that few fallibilists would be happy to require such a high improbability of error and that the Gettier problem below stands.)

Second, if there are errors in the favourable area, one can build Gettier cases. Suppose there is an error case in a favourable area of the photograph series. In that case, $S$ judges that some person at the front row is his friend; the photograph is quite clear but $S$ is mistaken. Now alter the case so that one person in the back is $S$’s friend. $S$’s belief is true in the modified case; it is also based on a reliable method and made in the favourable area. But it is not knowledge. In the addition series, a Gettier case can similarly be created by introducing mutually cancelling errors. (See Sosa, 1985, 240, Zagzebski, 1994, 69, Merricks, 1995, 845 for a similar argument.)

The Gettier problem arose for the simpler reliability requirement as well. Suppose that reliability across all the series is sufficient for knowledge (in conjunction with belief and truth). Then the Gettier cases above will be wrongly classified as knowledge; so the simple reliability requirement is not sufficient. Now if we move into a favourable area, but still merely require reliability in the favourable area, the same problem reappears. No requirement of reliability (or combination thereof) over natural ranges is sufficient for knowledge.  

35. As Sosa (1985, 239) puts it: “If a process is reliable but fallible, it can on occasion fail to operate properly. But that opens the possibility that a process yield a belief through improper operation where by lucky accident $B$ [the belief] turns out to be true anyhow.” Our reasoning shows that if one requires that the process operates properly, where properly is again less than infallibility, the problem reappears.

Sosa does not propose a solution to the problem in that paper, though infallibility is suggested (“What might be missing? Perhaps some closer connection between the belief and its truth? Perhaps these cannot be so independent as when they come together only by lucky accident, the way they do in our example of fallible memory.” p. 240) He concludes that the problem is not so much a problem for reliabilism as a theory of justification than for the idea that knowledge is justified true belief. The implication is thus that justification is not infallible, but that some other requirement on knowledge may be so.
4.2. The Infallibility constraint

4.2.4 Formal version of the argument

The argument above can be represented in a simple model. Let $W$ be a natural series of cases, such as the photograph series. Cases are centred worlds with a distinguished proposition. We write $K\alpha$ if $\alpha$ is a case of knowledge, $B\alpha$ if it is a case of belief, $T\alpha$ if it is a case of truth, and $E\alpha \equiv B\alpha \land \neg T\alpha$ if it is a case of error.\(^{36}\) Let $R$ be any condition that is satisfied by a case $\alpha$ in $W$ if error is avoided in most but not necessary all cases in $W$. We focus on a setting in which $W$ does include an error case. We build a case $\beta$ such that $R^*\beta \land B\beta \land T\beta$ but $\neg K\beta$, where $R^*$ is an epistemically natural extension of $R$.

First we create a duplicate model $W^* = \{<\alpha, 0>: \alpha \in W\} \cup \{<\alpha, 1>: \alpha \in W\}$. That is, for each case $\alpha$ in $W$, we have two cases $<\alpha, 0>$ and $<\alpha, 1>$ in $W^*$. Write them $\alpha_0$ and $\alpha_1$, for short. Let $p$ be the proposition that obtains in all and only $\alpha_1, \beta_1, \gamma_1, \ldots$ cases. Intuitively, we may think of $p$ as being the proposition that a certain distant coin has fallen heads. For each original case $\alpha$, we get a case in which the coin has landed heads ($\alpha_1$), and a case in which the coin has not landed heads ($\alpha_0$). Moreover, we change the subject’s beliefs as follows: for each original case $p\alpha$ in which she believed $p\alpha$, she believes $p\alpha \lor p$ in each of $\alpha_0$ and $\alpha_1$. The subject forms the belief on the same basis as in the $W$ series. She has no information at all with respect to $p$. She directly forms the disjunctive belief; she does not infer it from $p\alpha$.\(^{37}\) We claim that:

(A) For any $\alpha$, $K\alpha$ iff $K\alpha_0 \land K\alpha_1$.

Given the setup, if the subject knows $p\alpha$ in an original case $\alpha$, then she knows that $p\alpha \lor p$ in the duplicates $\alpha_0$ and $\alpha_1$. Switching to the disjunction cannot hurt. Moreover, if she does not know that $p\alpha$ in $\alpha$, then she does not know that $p\alpha \lor p$ in $\alpha_0$ and $\alpha_1$ either, for her epistemic position with respect to the disjunction is entirely dependent on her epistemic position with respect to $p\alpha$.

\(^{36}\) See 3.1.1 for definitions and p. 459 for a list of symbols.

\(^{37}\) As in Lehrer’s (1974, 19–20) clever reasoner case.
We also duplicate the relation $R$ by defining a relation $R^*$ such that for any $\alpha$, $R^*\alpha_0 \land R^*\alpha_1$ iff $R^*\alpha$. Our second premise is:

(B) $R^*$ is an epistemically natural extension of $R$ to $W^*$.

For any $\alpha$, $\alpha_1$ and $\alpha_0$ are cases in which the subject’s epistemic position is as in $\alpha$. Assuming that $R$ expresses, for $W$, some epistemically significant condition, that condition is satisfied in cases $\alpha_1$ and $\alpha_0$ iff it is satisfied in $\alpha$. So $R^*$ expresses for $W^*$ the epistemically significant condition that $W$ expresses for $R$.

(Note that if $R$ is formulated in terms of proportion of true belief cases, the definition implies that $R^*$ requires a higher proportion. For given the duplication procedure, if $W$ contained, for instance, 50 cases including 10 errors, then $W^*$ contains 100 cases but only 10 errors, since for any error case $\alpha$ we have introduce an error case $\alpha_0$ and a true belief case $\alpha_1$. This seems to me to be more a problem for fixed-proportion accounts of epistemically significant conditions than a problem for assumption (B). Note also that even if $R^*$ requires a higher proportion of true belief cases, it will remain strictly fallibilist if $R$ is.)

By assumption, there is some error case $\alpha$ in $W$. And by assumption as well, $R$ is satisfied in $\alpha$. Since $R$ holds in $\alpha$, $R^*$ holds in $\alpha_1$. But $\alpha_1$ is a case of true belief, since the proposition of $\alpha$ is $p \lor p$ and $p$ holds in all cases $\alpha_1, \beta_1, \ldots$. So we have $R^*\alpha_1 \land B\alpha_1 \land T\alpha_1$. But since $\alpha$ is a case of false belief, $\neg K\alpha$, and by assumption (A), $\neg K\alpha_1$. So $R^*$ is not sufficient, in conjunction with belief and truth, for knowledge. By assumption (B), the significant epistemic condition expressed by $R$ is not sufficient for knowledge either.

In intuitive terms, the pair $\alpha_0, \alpha_1$ is a Gettier situation: in $\alpha_1$, the subject has a true belief, but that is “just due” to the toss of the coin, because in $\alpha_0$ the subject would have formed a false belief.

The argument shows that if we have a reliability condition that is less than infallible in a series $W$, we can build a Gettier case involving a natural extension of that condition in a series $W^*$. This shows that the latter is not sufficient for knowledge; and by extension, that the original
Figure 4.1: Lottery cases and Gettier cases

Lottery cases

Gettier cases

- area without error
- area of error
- arbitrary distribution of error

condition was not genuinely sufficient either.\textsuperscript{38}

Visual representations

The schemas below may help. Figure 4.1 gives representations of lottery cases and Gettier cases. The circle represents a natural range of cases — a set $W$. In lottery cases, errors are arbitrarily spread in the range.

\textsuperscript{38} It may happen that $W$ is set up such that the reliability condition \textit{de facto} entails knowledge, though the natural extension of it does not. Suppose that we have an infallibility sphere $W_1$. Suppose that there is a reliability sphere $W_2$ around it ($W_1 \subseteq W_2$), such that any $\alpha \in W_2 \setminus W_1$ is a case of error. (That is, all the cases that are in the larger sphere but not in the smaller one are error cases.) In $W_2$, reliability \textit{and truth} is sufficient for being in $W_1$. So if $W_1$ is the knowledge sphere, it might look as if reliability \textit{and truth} in $W_2$ are sufficient for knowledge. But the natural extension of $W'_2$ of $W_2$, built as above, includes some true belief cases that satisfy the natural extension of the reliability condition but that fail to be knowledge. (Namely, all the cases $<\alpha, 1>$ where $\alpha \in W_2 \setminus W_1$.) This shows that even though \textit{de facto} the reliability condition is extensionally adequate in $W_2$, the “real” knowledge condition was being in $W_1$, not having a true belief in $W_2$. That is why the reliability condition will not be \textit{genuinely} sufficient, though it may be \textit{de facto} so.
Gettier cases are more structured. Take the fake barn case, for instance. In ordinary fields, Bernard is not mistaken about barns: thus most of our circle is white — this is why Barnard’s belief is justified or reliably based. But there are unfavourable conditions in which errors abound, such as the fake barn area in Bernard’s case. A Gettier case is a case of true belief within the unfavourable area. (On the schema, the white spot in the black circle.)

The foregoing argument against reliability requirements is illustrated in figure 4.2. The left circle represents some natural range $W$ in the original series. The subject is merely reliable, not infallible, in the range: so we have an error area (the black circle). Now we construct a Gettier case as follows. We duplicate the area according to the value of an unrelated proposition $p$, and we change the subject’s beliefs to $p_\alpha \vee p$ in each belief case $\alpha$. We thus get two “circles”, one in which there is no error at

---

all (since \( p \) is true all over), and one in which there is error just where there was error in \( W \). The two circles are superposed: the \( p \)-version and not-\( p \) version of each original case are likewise situated in epistemically relevant respects. Where the original error area was, we now have an area in which the difference between truth and error is just a matter of luck, as in a lottery. The true beliefs in that area are Gettier cases against the claim that being within that circle of reliability is sufficient, in conjunction with truth, for knowledge. Gettier cases are in effect lotteries in unfavourable areas.

The infallible method solution to be defended here is illustrated by figure 4.3. Knowledge requires infallibility over a natural range. No error case should be included in the relevant range, otherwise Gettier cases are again possible.
4.2.5 Conclusion

Reliability requirements allow a discrepancy between belief and truth that generates Gettier cases. Reliability is compatible with error in the relevant area; error in the area is compatible with accidental truth in the area, and accidental truth in the area is a Gettier case.

To avoid Gettier problems, a modal requirement on knowledge should require the impossibility of errors based on some method in a relevant range. In some of the photograph cases, the photographs and the subject should be such that error was simply impossible in the case; more precisely, the case was such that it belongs to a natural range of cases in which error is excluded. And similarly for the series of additions.40

Method infallibility is at least necessary for knowledge. For some relevant notion of possibility, we have:

**Moderate Method infallibilism**  $S$ knows that $p$ only if $S$’s belief is based on a method such that all possible cases involving the same method are cases of true belief.

(I call this version “moderate” to contrast it with the “radical” one according to which method infallibility is necessary and sufficient for knowledge. The radical one is advocated in sec. 4.4.) Formally:

**Moderate Method infallibilism**  For any $\alpha$, $K\alpha$ only if $\exists m (Bm\alpha \land \forall \beta (\alpha R\beta \rightarrow \neg E_m\beta))$, where

- $Bm\alpha$ iff $S_\alpha$ believes $p_\alpha$ on the basis of $m$ in $\alpha$,
- $R$ is an accessibility relation that selects a natural range of cases around $\alpha$,
- $E_m\beta$ iff in $\beta$ the subject has a false belief based on $m$: $Bm\beta \land \neg T\beta$.

Models are given in ch. 5 by introducing a formal representation of methods.

---

40. A similar view is defended by Armstrong (1968, 190), Dretske (1971, 11–2), and Pritchard (2007, 281–4) in reply to Greco (2007, 112–4), though the three proposals are proposition-centred.
4.3 Objections to necessity

Requiring infallibility for knowledge raises the worry that knowledge is unattainable. Using the notion of method raises the question of how methods are individuated. Each problem threatens moderate method infallibilism. The problems are not independent: whether one’s methods are infallible depends on what one’s methods are. My answer to both relies on the same methodological point: we have to use our intuitions about knowledge to find out what the relevant notions of method and possibility are. The notion of method is not antecedently or independently given; nor can we assume that it is characterisable or applicable without using the concept of knowledge or related concepts. Similarly, we have to use intuitions about knowledge to find what notion of possibility is relevant to method infallibility. We cannot assume at the outset that it is identified with physical or metaphysical necessity, for instance.\(^{41}\)

In this section I defend the methodology and present some of its results.

4.3.1 The generality problem

The generality problem has originally been stated for Goldman’s process reliabilism (Goldman, 1979, 12; Conee and Feldman, 1998, 1). Any particular process belongs to vast number of types. Only process types can be said to be reliable. A particular process is not reliable or unreliable: either it has produced a true belief, or it has not, but since it is not repeatable, we cannot ascribe it a frequency or tendency of success (Goldman, 1979, 11). So in claiming that a belief is justified iff the process that led to it was reliable, the reliabilist must single out a relevant process type. The generality problem is to state which one.\(^{42}\)

---

41. A similar answer to the generality problem has been advocated by Williamson (2000, 100; 2009a, 19–20).
The problem got its name from the fact that various process types have various levels of generality. In Goldman’s (1979, 12) formulation, the problem is to tell “the degree of generality of the process-types in question”. But of course one should not assume that the “degrees” of generality of process-types form a scale: types may cut across each other. Nor should one assume that there is a unique “degree” of generality that all relevant process-types share: epistemic evaluation may invoke both very specific processes (in perception, say) and very general ones (in deduction, say). Goldman’s idea that particular processes do not have a frequency or propensity of success can also be challenged (Comesaña, 2006, 28–9): if particular processes are individuals, they may exist in counterfactual situations, and have a frequency of success over the modal space. (Compare with a particular atom’s propensity to go left at a certain time.)

In our framework, we restate the problem as follows. Methods are epistemically relevant classes of belief cases. We may assume that each belief case involves at least one method, the method of believing. But we do not assume that a particular belief case involves only one method. Method infallibility states that knowledge in a case requires that one of the methods of the case is an infallible one, not that the method of the case is an infallible one. So we do not have the problem of saying which method is the one used in a given case. But we do have the problem of saying which classes of belief cases are epistemically relevant, that is:

**The generality problem** Which classes of belief cases are equivalent classes of some single method?

---

Armstrong (1973, 166-7) seems to have barely missed the problem: “Laws of nature are connections between things […] of certain sorts. But the suggested connection between belief-state and situation which makes the belief true is a connection between particular states of affairs: between singulars. The reply to this is that the belief-state is an instance of a certain sort of thing […]. Equally, the situation is an instance of a certain sort of thing.” The ensuing discussion rules out general sorts, so Armstrong seems aware that a particular is an instance of many sorts; but he does not give a clear rule to select relevant ones. If anything, the text (p.167) suggests that any property (in the sparse sense, I assume) of the states of affairs will do.

43. As we pointed out in the introduction, p. 5.
4.3. Objections to necessity

That is, under what conditions a class of belief cases $C$ is such that there is a method $m$ such that for all cases $\alpha$, $\alpha$ is a case of belief based on $m$ if and only if $\alpha \in C$?

The generality problem is thus to provide conditions under which a class of cases constitutes (the extension of) a method. A complete answer would provide conditions for some classes; a full answer would provide conditions for all.

The generality problem can be understood in two ways, as Kappel (2006, 528–9) notes:

The criterial problem What are the criteria that allows us to tell that a class of cases constitutes a method?

The grounding problem What makes it the case that a class of cases constitutes a method?

The first problem is epistemic: an answer to it would put us in position to know which classes constitute methods. The second problem is metaphysical: an answer to it would tell us what has to be the case for a class to constitute a method, but that may not put us in a position to tell which classes fill the bill. 44

The difference between the two questions is illustrated by contrasting Schmitt’s and Alston’s answers to the generality problem for process reliabilism. On Schmitt’s (1992, 141–2) view, “we have intuitions about which processes are relevant”. We have common-sense concepts of the relevant types. That would ensure that we can tell what the types are, but it would not tell us why these types are more important or relevant than others. 45 Alston’s (1995, sec. VI) view is that the relevant type of process is the one realised by the psychological mechanism that was causally operative in forming the belief. He defends a psychological realism that posits such mechanisms and their unicity. That tells us what

44. Kappel (2006, 528–9) calls the first problem that of finding criteria of relevance and the second that of offering a theory of determination. Conee and Feldman (1998, 3) formulate the problem in terms of “identification”, which can be read in both ways: “A solution identifies the type whose reliability determines whether a process token yields justification.”

45. See also the view attributed to Goldman by Conee and Feldman (1998, 7).
is special about the types thus singled out. But that may not enable us to
tell whether two cases instantiate a same process of a relevant type.

Now each question is associated with an implicit constraint on ac-
ceptable answers. For the criterial question, answers like *when it is an*
*epistemically relevant class* are not excluded. We are expected to give cri-
teria that can be applied without applying the concept of knowledge (Conee
and Feldman, 1998, 4). For the grounding question, answers like *when its infallibility entails that one knows* are excluded. Methods cannot earn
their metaphysical living simply by their role in epistemology (Conee
and Feldman, 2004a, 5).

Now why accept the constraints? I think each has a reasonable moti-
vation:

**Testability** We want to test method infallibilism against ordinary judg-
ments about knowledge.

**Reducibility** We want an account of knowledge that reduces it to more
basic notions. Knowledge is not fundamental. It is made out of more
basic things and properties, like belief, causal relations, probability,
and so on.

The first desideratum motivates the criterial problem. The second mo-
tivates the grounding problem. What I want to argue here is that *both desiderata can be satisfied without answering the generality problems*. While an
account of methods that would satisfy the implicit constraints of the gen-
erality problems *would* fulfil the desiderata, an account may fulfil them
without answering the generality problems.

First I will say why I doubt that the epistemic generality problem can
be solved (4.3.2). Then I will say (4.3.3) why this does not prevent method
infallibilism from being testable, nor from being reductive.

### 4.3.2 The concept of knowledge is primitive

The concept of knowledge is a primitive among our ordinary con-
cepts. We acquire it very early, and not through the exposition of some
4.3. Objections to necessity

rudimentary theory. (Contrast it with the concepts of wage or election, for instance.) We seem to apply it through simple heuristics such as:

1. If someone give a true answer to a question, they know it.
2. If I know that \( p \) and told someone that \( p \), they know that \( p \).
3. If you and I see the same scene, you know everything I know about it.
4. If not-\( p \) but you believe \( p \), you do not know that \( p \).
5. If I told to you a lie that \( p \), you do not know that \( p \).

With the exception of (4), the heuristics need involve applications of the concepts of truth, belief nor a self-application of the concept of knowledge. If I believe \( p \) and (I believe that) you answered \( p \) to a question, I can directly ascribe you knowledge that \( p \) without ascribing belief to you or myself or believing that \( p \) is true.\(^{46}\) If I believe that \( p \) and (I believe that) I told you that \( p \), I can directly ascribe you knowledge that \( p \). And for any perceptual belief \( p \) I have about the scene we are both watching, I may ascribe you knowledge that \( p \). The heuristics tend to ascribe knowledge in cases of true belief and deny it in cases of non-belief, false belief, and belief based on a false belief. They are fairly reliable, provided that accidental true belief is rare.\(^{47}\)

We undoubtedly supplement such heuristics with conceptions of sources of beliefs such as perception, testimony or memory. But there are two important points to note. First, these conceptions may be themselves based on the primitive concept of knowledge. Seeing that \( p \) may be prior to having a visual perception as if \( p \). Second, these conceptions need only provide more refined heuristics. That is, they may type beliefs in such

\(^{46}\) Provided that believing that \( p \) is true is more than simply believing \( p \) — the former, but not the latter, requires entertaining the concept of truth. “I believe that” is in brackets in the antecedent because it may not even be needed that I believe that you answered \( p \). The fact that you did so may directly cause my belief that you know without intermediate belief.

\(^{47}\) An important point to note is that such heuristics focus on beliefs/propositions that are salient to us, not all beliefs one may have. We typically focus on informative beliefs/propositions. The reliability of the heuristics may depend on this.
ways that beliefs in a type merely tend to be knowledge. So we cannot assume that our ordinary conceptions of sources are adequate conceptions of methods.

If the concept of knowledge is primitive in that way, there is no reason to expect that it is analysable in terms of other ordinary concepts. Even if it was, there would be no reason to expect that we are able to apply the relevant concepts without applying the concept of knowledge itself. For this reason, I am sceptical of accounts of methods that would assume that methods are things of which we have ordinary concepts (along the lines of Goldman, 1979, 9–10 or Schmitt, 1992, 141–2). 48

In all likelihood, the epistemic generality problem cannot be answered with ordinary resources that are independent of the concept of knowledge and its applications. Can it be answered with scientific resources that are so independent, as Alston (1995, 11) thinks we eventually can? 49 I do not know. But suppose that we cannot build such a science of methods without using the ordinary concept of knowledge itself. That would imply that the epistemic generality problem cannot be answered. Let us assume that it cannot: does that imply that methods infallibilism is not testable and that knowledge is irreducible?

48. Incidentally, the same considerations suggest that there is a limit on our ability to take a higher-order perspective on our ordinary knowledge — a central tenet of Sosa’s view (Sosa, 1991, 1997b). Our beliefs about first-order knowledge may either involve knowledge-involving concepts (e.g. seeing that p, having an adequately grounded belief) or the rough concepts we use in heuristics of knowledge ascriptions (e.g. having a visual experience). In the former case, they will be fine-grained enough to capture one’s epistemic situation precisely, but they will not provide independent concepts of methods. In the later, they may provide independent concepts, but the concepts will not discriminate between knowledge situations and some similar non-knowledge situations such as Gettier cases.

49. To clarify my view of Alston’s reply to the Generality problem: by saying that there are real psychological mechanisms that determine which types are relevant, he gives an answer to the metaphysical question without giving an answer to the epistemic one. But he implies that once psychologists have done their work, we will be able to answer the epistemic question by using their concepts of the mechanisms.
4.3. Objections to necessity

4.3.3 Methods can be more basic than knowledge

We do not assume a prior and independent notion of methods (4.1.2). Rather, we propose to introduce them via method infallibilism itself: 50

Initial Method theory Methods are natural properties of belief cases the infallibility of which is necessary (and perhaps sufficient) for knowledge.

Methods are thus characterised as properties of belief that make method infallibilism come out true. This does not imply that method infallibilism is true by definition. The description may fail to refer. The only properties that satisfy it may be much less natural than close properties which do not, in which case there will be a lot of pressure to theorise in terms of the latter and drop method infallibilism. If several properties satisfy the description, the notion of method will pick the most natural ones, or we will end up with a family of notions of methods — each of which would yield a different notion of knowledge. 51

Thus if anything, Method infallibilism is a definition of methods rather than a definition of knowledge. It gives us a partial theory of what methods are. It also gives us partial criteria for methods, but ones that rely on our judgements about knowledge:

Criterion 1 If α is a case of knowledge and β a case of error, then:
    either there is a method that α instantiates but not β,
    or β is not in the natural range of α.

Criterion 2 If α is not a case of knowledge, then:
    for each method instantiated in α, there is an error case β in the natural range of α that instantiates the same method.

As the two options show, the characterisation sets a problem with two variables, methods and possibility. Given a knowledge case, we can

50. The approach we adopt with respect to the notion of methods is akin to the one defended by Lewis (1970b) with respect to theoretical notions in general.
51. The fact that we rely on our judgements about knowledge to figure out what methods are does not guarantee that we will not end up with a pluralistic theory of knowledge. We may find out that what we call knowledge is more naturally seen as a range of of properties.
either individuate methods broadly and restrict the range to keep error cases at bay, or we can allow a wide range and individuate methods more narrowly. One extreme, as we pointed out (sec. 4.1.4), is a variant of Williamson’s similar-cases safety in which the range is unrestricted.

The task is to find a package of a notion of method and a notion of possibility that is natural and fits our judgements about knowledge. There are several ways to constrain the notions. One is to specify their structural properties. We make a first step in that direction in ch. 5. Another is to directly impose constraints that appear plausible or natural. Here are two candidates:

**Methods are history** if $\alpha$ and $\beta$ differ only with respect to their future, they do not differ in methods used. \(^{52}\)

**Space-time regions are natural** if $\alpha$ and $\beta$ are located within a sufficiently small space-time region of the same world, they are in each other’s relevant natural range. \(^{53}\)

---

52. **Revision note.** Ernest Sosa points out that it is not clear why, if methods can involve parts of one’s spatial environments, they could not similarly be extended towards one’s future. He also argues that if methods are not so extended, the future will not be knowable — but that seems prima facie false.

The sceptical consequences of the history constraint are alleviated by the restriction of the relevant space of possibilities. If the possibilities relevant to evaluate knowledge contain at most the physically possible ones, not any future will be possible given a certain history.

I do not have a fully satisfactory answer to the first point. Perhaps a better methodology would be to start with strong internalist assumptions on the individuation of methods (narrow and a “methods are present” constraint) and see where they lead, weakening them as problems are encountered. Or the opposite. There is no clear justification for the present strategy, which treats the “method are history” constraint as an assumption and the narrowness constraint as an open question.

53. **Revision note.** Ernest Sosa objected a pair of cases. (a) one is spatially close to various misleading objects (e.g. fake barns), but these are located in an area in which there is no possibility at all for subjects to go. The error cases are spatially close but not relevant. (b) one is watching at a distant scene from a TV screen. Error cases are spatially distant but relevant.

These particular cases may be disposed with in the present approach. In (a), while there are cases of error with a nearby location, these will not be possible on the relevant notion of possibility. In (b), while the scene watched is distant, the case of error is located where the subject is, so it is in fact a spatially close case.

Further cases may be brought against the principle, however. (a) Suppose that a given area is stably divided in two almost causally independent regions, and that
4.3. Objections to necessity

A further one is to enquire about further constraints on ranges of cases. We do some of this in the following section (4.3.4) and in appendix E. Finally, the notions are further determined by their relations with other notions than knowledge. In ch. 7, we argue that the relevant notion of possibility is the one that is at play in modals and counterfactuals.

Method infallibilism is not tested by making pre-theoretical intuitive judgements on methods and their infallibility and comparing them to our intuitive judgements on knowledge. But it is tested by whether a suitably natural theory of methods and of their infallibility can be devised that matches what we independently know about knowledge. Given that any modal approach needs a notion of errors based on the same method (4.1) and that fallibilist modal accounts face Gettier problems (4.1.1), the unavailability of such a theory would mean that modal accounts of knowledge should be given up.\(^{54}\) Methods infallibilism is falsifiable, hence testable.

Now let us turn to the metaphysical side. Suppose we build a notion of method \textit{via} a theory which essentially uses the concept of knowledge. Does this mean that knowledge cannot be \textit{reduced} to methods and infallibility? Obviously not. If we pick out atoms as being the elementary constituents of ordinary objects, it does not follow that atoms are somehow ontologically dependent on ordinary objects or that ordinary objects cannot be reduced to atoms. Similarly the fact that we only get a grasp of methods through their relations with knowledge does not prevent meth-

---

\(^{54}\) Sosa's (2007, 28–9) recent abandonment of safety in favour of an explanationist account can be taken as an endorsement of this view.
ods from being ontologically more basic than knowledge is. To use an older terminology, even though knowledge is the first in the order of knowledge, methods may be the first in the order of things.

I am particularly inclined to see methods as a reductive basis for knowledge if on the one hand, methods are seen as relations between possibilities and beliefs (as in the formal model developed in chap. 7) and, on the other hand, knowledge is taken to be a concept that is evaluative or normative in a way that sets it apart from non-normative concepts like that of *true belief*.

Thus there is no obstacle to claiming that knowledge is made out of methods, even though our initial theory of methods involves the concept of knowledge.

### 4.3.4 Application: fine-grained, non-luminous methods

Let me apply the methodology laid out in the foregoing section. We assume that method infallibility is necessary for knowledge:

**Moderate Method infallibilism** For any case $α$, if $S_α$ knows that $p_α$ in $α$, then there is a method $m$ such that in no case $β$ accessible from $α$ does $S_β$ believes $p_β$ on the basis of $m$ and $p_β$ is false.

We assume that accessible cases include all actual cases within a sufficiently small space-time region:

**Space-time regions are natural** if $α$ and $β$ are located within a sufficiently small space-time region of the same world, $β$ is accessible from $α$.

Take a sufficiently small space-time region to be of a few hours and an acre, for instance. Finally, we help ourselves with intuitive judgements about knowledge in such cases. Here are a few lessons we may draw from such settings about how methods are individuated.

**How fine-grained are methods?**

Methods can be very fine-grained. Consider the following series of cases:
Magoo and the leaf  Magoo has extremely poor sight. Anything he sees at half arm length or more is quite blurry to him, and both colours and shapes are distorted. He needs to take things very close to his eyes to identify them. Magoo is also extremely bad at recognising tree leaves, even from up close. Though he has tried to learn to identify leaves of common trees, he is liable to confuse the leaves of elms, beeches, poplars, and so on. Nevertheless, for some reason, he has perfectly memorised the shape of (common) oak leaves. (For they remind him of jigsaw puzzle pieces.)

Now Magoo is spending a few hours in the lab, being shown various leaves from various distances and for various durations. He makes more or less confident judgements about their species. He often mistakenly takes himself to have rightly identified a leaf. In one case ($\alpha$) he is shown an oak leaf from extremely near and is given a full minute to examine it. He declares it to be an oak leaf with more confidence than any of his other judgements.

Magoo knows in $\alpha$. Yet case $\alpha$ is in a range that includes errors made on the basis of the same sensory modality. Some of these errors also bear on the species of a leaf seen from very up close. Other errors also bear of the species of an oak leaf but seen from a further distance. Thus there has to be a method instantiated in $\alpha$ that did not encompass cases of identification of the same life at greater distances nor cases of identification of leaves of other species. Methods can be very fine-grained.

Note that I am only using the example to show that methods can be quite fine-grained. I am not deriving individuation conditions from it — not even implicitly. First, the fact that Magoo has a distinct level of confidence in $\alpha$ plays a role. If Magoo’s confidence was equal in all cases, it would be less clear that he knew in case $\alpha$. Fine-grained individuation may require that the subject is somehow sensitive to the difference in methods. So one should not assume that methods are always as fine-grained as in that series.  

55. See Sosa’s (1999a, 379) remarks that knowledge requires should be somehow “guided” by the conditions under which one’s judgement tracks the truth.
Second, it is clear that the sensory modality is relevant, but nothing in the case allows one to systematically individuate some methods by the kinds of percepts involved. For instance, the idea that there is always a method that groups only the cases that involve the roughly same visual experience probably leads to disastrous consequences. Consider for instance a series based on our Wedding Ring case (see p. 162):

**Wedding Ring Series** As in the original case, Ringo would inevitably wear a ring if married, but many of his married colleagues do not. The foreign client successively meets three different-looking reception clerks without rings, including Ringo, and forms about each the belief that they are not married.

The foreign client does not know. But if his belief instantiates a percept-individuated method, and if method infallibility is to explain failure of knowledge in the case, then one needs either accessible cases in which one of Ringo’s colleagues looks like him, or in which Ringo would be ringless though married. It seems more natural to say that the difference in looks between Ringo and his colleagues are irrelevant to any method the foreign client’s belief is based on. So it is not the case that for any belief based on a sensory modality, we get a method that is individuated by the corresponding percept.

A dialectical point worth stressing here is that *reliability requirements commit one to such fine-grained methods too*. Many epistemologists take knowledge to require belief based on a reliable process (Goldman, 1986, 43) or an epistemic virtue (Sosa, 1991, 271, Greco, 1999, 287, Pritchard, 2010). They do not commit themselves to the view that the relevant processes or virtues are infallible. Yet cases like the Magoo series would lead them to adopt equally fine-grained conceptions of the relevant processes and virtues. Magoo’s tree leaf identification is not only fallible, it is utterly unreliable. His vision is not virtuous; if anything, it is vicious. Even his vision from up close need not be virtuous. We may assume that even when looking at things from up close, their shapes appear to him

---

56. Lewis (1996, 533) endorses this idea.
as moving. But still, despite the movement, he is able to spot the oak leaf. Despite all this Magoo knows that he is seeing an oak leaf. So while reliability requirements may afford fairly general methods in the series of additions or in the series of photographs (p. 173), they cannot do so in cases like Magoo’s.  

Are methods luminous?

The Magoo series suggested that a subject needs to be sensitive to relevant differences between cases, in his level of confidence for instance. Should relevant differences something the subject is in position to know? Should it be that:

Method luminosity

If $\alpha$ is a case of belief, and $\beta$ an accessible case not involving some method that $\alpha$ involves, $S_\alpha$ in $\alpha$ is in a position to know that $\beta$ does not involve a method that $\alpha$ involves.  

The series of additions and series of photographs cases (p. 173) show that luminosity is false. At any point in the series, the subject may believe that his beliefs in each part of the series are based on the very same methods. This would not prevent him from knowing in the favourable cases in spite of his error in the unfavourable one. In the photographs series, there may be a requirement that the man be somehow sensitive to the different levels of graininess; but that does imply that he is in position to know that he is using different methods on better photographs. In the addition series, it does not appear necessary at all that the subject be in position to know that his attention is better in the knowledge case than in the error case.

This is related to our earlier claim that the ordinary notions of bases of beliefs may be too coarse to give us adequate notions of methods (sec. 4.3.2).

57. A reply that the defender of broadly-individuated virtues (or cognitive processes) may make is to stress that virtues are not subject-centred. Perhaps one can argue that what Magoo is using is the faculty of vision, and that vision in other people than him is reliable. So Magoo’s visual beliefs are reliable after all. I do not fine the line plausible, but I will not discuss it further here.

58. See Williamson (2000, 95) for the notion of being in a position to know.
Are methods narrow?

Methods are narrow iff any two cases that are internally alike do not differ by the methods used; they are broad otherwise (Williamson, 2000, 66). Being internally alike is relative to a notion of what counts as the inner. For simplicity we focus on the physical inner: two cases are internally alike iff the internal physical state of the subject is the same. Williamson (2000, 66–72) has put forward a way of building cases to show that various mental states are broad. (Williamson’s aim is to show that they are prime, that is, not equivalent to the conjunction of a narrow and an environmental condition. But I focus on broadness here.) Interestingly, under the limited assumptions we have made about accessibility, such cases cannot be used to show that methods are broad.

Consider the following pair of cases, abstracted from Williamson’s (2000, 72) cases to argue for the primeness of knowledge:

$\alpha$ Smith tells one that the election was rigged, he is trustworthy, and one trusts him.

$\gamma$ Smith tells one that the election was rigged, he is not trustworthy, and one trusts him.

In case $\alpha$, one knows that the election was rigged, but in case $\beta$, one does not, even though the internal physical state of the subject is the same. In Williamson’s setting, the cases may belong to any possible worlds, so we can arguably find such a pair. Together, they show that knowledge is broad.

Can we use such pairs to prove that methods are broad? We have made only one assumption about accessibility: that cases taking place in roughly the same space-time region of the actual world as the target case are accessible. Thus our pair of cases will have to take place roughly at the same time. To avoid controversy over subject-centring (see ch. E), it is better if they involve the same subject. This raises the question whether a subject can be in the same total internal physical state at two different points in time in the same world. We assume that this is so if the subject is put to (brief) sleep and woken up, each time with is memories of the
4.3. Objections to necessity

previous case erased.\footnote{As in sleeping beauty cases (Elga, 2000).} We also have to use tensed propositions, to allow Smith to tell a truth in one case and lie in the other while the two cases are in the same world. We get the pair:

\begin{itemize}
\item $\alpha'$ Smith tells $S$ that it is raining, he is trustworthy, and $S$ trusts him.
\item $\gamma'$ Thirty minutes later, $S$ has no memory of the previous episode. Smith again tells $S$ that it is raining, but he is now not trustworthy, and $S$ trusts him.
\end{itemize}

It seems clear that the subject does not know in either case. We get the same judgement with perceptual cases:

\begin{itemize}
\item $\alpha''$ At $t_0$, $S$ is presented with a orange and forms the belief that it is an orange.
\item $\gamma''$ At $t_1$, $S$ is presented with a wax replica and forms the belief that it is an orange.
\end{itemize}

Again it seems clear that $\alpha''$ is not a case of knowledge. In both cases one fails to know precisely because one is going to encounter an internally alike case in which one is wrong in the next hour. These cases suggest the following principle:

\textbf{Close hallucination prevents knowledge} For any $\alpha, \beta$, if $\alpha$ and $\beta$ take place in the same space-time region of the same world and are internally alike (and perhaps involve the same subject), if $\beta$ is an error case then $\alpha$ is not a knowledge case.

The principle is consistent with, but does not entail, the claim that methods are narrow:

\textbf{Narrow methods} For any cases $\alpha, \beta$, if $\alpha$ and $\beta$ are internally alike then a method is used in $\alpha$ iff it is used in $\beta$.

However, while I cannot find a clear counter-example to the principle, not all judgements are clearly in its favour. If the testimony pair $\alpha' - \gamma'$, if $\gamma'$ is altered so that some other person (Brown) gives the unreliable testimony, the case seems to me less clear. (Whether Brown looks or
sounds like Smith or is even seen or heard does not seem clearly relevant to me.) In Unger’s series of additions (p. 173), it is hard to get a pair of cases in the series in which one would be in the same internal physical state while being wrong in one case and right in the other. But at least it does not seem at all required that there is anything phenomenologically special in the knowledge cases of the series. Finally, if we alter the orange case $\alpha'' - \gamma''$ so that in $\gamma''$ one does not see a wax orange but dreams of seeing an orange, it is less clear that one does not know in $\alpha''$. But it also presumably not possible that one is in the same internal physical state, and perhaps not even in a belief state.\(^{60}\)

To investigate further, one would need to consider cases that are more spatially and temporally distant. Consider in particular:

**Two Naturalists** Alice is exploring an island of the Atlantic in search of bird species. She spots a pair of birds of the same species, observes them well. She spots a third one that looks in all respects like the two first. She has the opportunity to observe each fully, and judges the third to be from the same species. She is right.

Pamela explores an island of the Pacific and has the very same bird experiences. In Pamela’s case, however, the third bird is not of the same species. It is of a related species that has branched away from the first a million years ago. One can only distinguish them by the colour of their eggs. In Alice’s case, the third bird is of the same species.

Alice’s birds are of the same species as Pamela’s first two. There was no chance for Alice to spot a bird of the cousin species on her island. The cousins species has never spread to the Atlantic.

A variant would put the naturalists apart in time rather than space.\(^{61}\)

Such cases raise thorny issues. First, when told the Alice part of the story, we readily ascribe knowledge, but in view of Pamela’s story, it is far less

---

60. See Sosa (2007, ch. 1) for further discussion.

61. Further variants involving a third, travelling naturalist and communication between the two naturalists raise interesting issues as well.
4.3. Objections to necessity

clear that Alice knows that her third bird is of the same species. Second, suppose we say that Alice knows. By method infallibility, it follows that Pamela’s case is either different in method or a non-accessible case. There are several ways to go here. First, one can use subject-centring: Pamela’s error is irrelevant to Alice’s knowledge, because it involves a different subject (ch. E). Second, one can use the accessibility relation: one could say that cases distant from Alice’s case in space and time are not genuine alternative possibilities for Alice’s case. Third, one can say that Alice’s case involves a method that Pamela’s case does not. For instance, the method could be characterised as identifying a bird species on the basis of features that are indicative of it in the island one is exploring. The method is broad: Pamela is not using it, even though her internal physical state is the same. But it is also such that any subject in the same internal physical state on Alice’s island is using it. It is thus compatible with the Close Hallucination principle stated above. If we reject subject-centring (as we advocate ch. E) and if we want to count all actual cases as possible in the relevant sense (as we advocate in ch. 7), the third option is mandatory.

Let me stop the preliminary investigation here. Since the individuation of methods is a multiple variable problem, it is better approached in a coherentist manner, by a parallel examination of constraints on accessibility. So far we can take stock of the following claims: (1) methods can be very fine grained, (2) methods need not be luminous, (3) in a number of cases (e.g. perception) methods need to be such that at least within the same spatio-temporal area internally alike cases cannot involve different methods.

---

62. The intuition that she does is perhaps more robust in the time variant.

63. Revision note. As Timothy Williamson pointed out, (3) plausibly leads to an internal individuation of methods. Take a series of internally alike cases such that each is sufficiently spatio-temporally close to the other. By (3), any two pair involve the same methods; hence all cases in the series do. Avoiding this problem may require a shift to a method-similarity approach (see sec. 4.3.6 above).
4.3.5 Scepticism and possibility

If knowledge requires a method that could not yield errors, it seems that we know next to nothing. Many would rather give up infallibility than endorse scepticism (Lewis, 1996, 550). Here I will argue that infallibilism does not entail scepticism. I do not endorse the anti-sceptical view either: a fuller discussion of the issue must await ch. 7.

My answer to sceptical worries can be gathered from my answer to the generality problem (4.3.3). Methods infallibility requires that error on the basis of some method was in some sense impossible. But the relevant notions of method and possibility are not given beforehand. Rather, we should rely on our judgements about knowledge to identify them. On this approach, the generality problem and infallibility-based sceptical worries are two faces of the same coin.

To ease the discussion, I make the following assumption:

**Actual cases are possible** For any cases $\alpha, \beta$, if $w_\beta = w_\alpha$ then $\beta$ is in the relevant range of $\alpha$.

The assumption is dialectically acceptable, since it can only increase the range of relevant cases, and thus it can only play in the sceptic’s favour. As the Two Naturalists case (p. 198) show, this is a strong constraint. Unless subject-centring is adopted, errors made in distant places and times become relevant.

Three types of cases suggest that method infallibilism leads to scepticism: radical sceptical scenarios, quantum-mechanical cases and ordinary cases. We discuss them in turn.

**Radical scenarios**

Here is the familiar line of reasoning of the Cartesian sceptic:

Take any case $\alpha$ of perceptual belief. There is a possible case $\beta$ in which one is in the same internal physical state and one’s belief is false. In $\beta$, one has the same experience as in $\alpha$. So any method used in $\alpha$ is fallible.
The argument crucially relies on two premises. First, it is assumed that methods are narrow, so that the internal likeness between $\alpha$ and $\beta$ is sufficient for their methods to be the same. Second, it is assumed that any case $\beta$ which results from the recombination of the internal physical state of the subject in $\alpha$ with another environment is possible. Clearly not all such cases are going to be physically possible. So the implicit claim is that knowledge requires at least metaphysically infallible narrow methods.

The assumption that methods are narrow is very contentious. But we can dispose of it by switching to Humean scepticism. Instead of narrow methods, we assume the much more plausible idea that the methods of a belief supervene on that belief’s past:

**Methods are history (MH)** If $\alpha$ and $\beta$ differ only with respect to their future, a method is used in $\alpha$ iff it is used in $\beta$.

Now we have an analogue to the Cartesian sceptically argument for knowledge of the future:

Take any case $\alpha$ of belief about the future. There is a possible case of belief $\beta$ in which the history is the same as in $\alpha$ but the belief false. By (MH), any method used in $\alpha$ is used in $\beta$. So the method used in $\alpha$ is fallible.

Given method infallibilism, (MH) and the claim that cases with the same past but a future as different as one likes are possible, one cannot know anything about the future. (Not even that the world will not suddenly be full of singing rhinoceroses.) Which is already bad enough, and which will plausibly spread to scepticism about many past claims too.

64. See Williamson (2000, 73).
65. See e.g. Williamson’s (2000, 174–7) criticism of the related phenomenal conception of evidence.
66. where $\alpha$ and $\beta$ differ only with respect to their future iff: the history of $w_\alpha$ matches the history of $w_\beta$ up to $t_\alpha$, $t_\beta = t_\alpha$, $S_\beta = S_\alpha$, $p_\beta = p_\alpha$.
67. For present purposes a belief is about the future in a case iff the history of the case does not entail its truth or falsity.
68. I put a ball in a box. I infer that it will still be there in one hour. By Humean scepticism, I do not know. The hour passes, and I still believe that the ball is there, on the same basis. Can I suddenly acquire knowledge that the ball is there now, simply because that fact is now part of history? That seems implausible.
The non-sceptical method infallibilist should reject the possibility claim. Not every metaphysically possible error makes error possible in the relevant sense for infallibility. Infallibility relies on a restricted notion of possibility. 69

Quantum-mechanical lotteries

Suppose the infallibilist settles for nomic possibility, what is possible given the laws of nature (Armstrong, 1973, 166). Now if certain views of quantum mechanics are right, just about any proposition we believe about the future has a non-zero objective chance of happening (Hawthorne and Lasonen-Aarnio, 2009). If we accept:

**Chance entails nomic possibility** If there is an objective non-zero chance that \( \neg p \), \( \neg p \) is nomically possible.

we can restate the Humean argument for nomic possibility. The idea that objective chance entails nomic possibility is not entirely uncontroversial. It is false concerning statistical notions of objective chance in deterministic settings. But if objective chance derives from indeterministic laws, the principle appears plausible. 70

Interestingly, the argument may be used to generate scepticism about what is currently perceived. Suppose I form at \( t \) the (tensed) belief that there is an apple on the table. Assume that the belief refers to the present time, that is the time at which the belief is formed. The belief is true if there is an apple on the table at \( t \). However, perception takes time; one’s perceptual state and belief are based on how the table was slightly before \( t \). Thus one could argue that ordinary perceptual cases are analogous to the ball case. One way to reject the argument is to claim that perceptual beliefs do not refer to the present time but rather to the time at which the situation perceived takes place. (See Récanati (2007, ch. 16, esp. 135) for a view of perceptual content that allows this.)

69. An early version of the idea is found in Armstrong (1968, 189): “If somebody had a true belief that logically could not be false, that would certainly be knowledge. But we have argued that this is logically impossible. But suppose that he had a true belief that *empirically* could not be false? A belief that *empirically* could not be false? A belief that empirically could not be false would be an absolutely reliable belief. Would it not give everything that was required for knowledge?” “Empirical” possibility is what we would call physical or nomic possibility. (The proposal is belief-centred, though, and thus violates the method constraint.)

70. Lewis’s Humean account of laws leads him to claim that some non-zero chance outcomes are nomically impossible (Lewis, 1994, 486–488). This (counter-intuitive) feature of the view is a consequence of the Humean reduction of laws to overall distributions
4.3. Objections to necessity

Widespread quantum indeterminacy is a surprising discovery. It is not expected on our commonsense conception of the world. It is not obvious that our commonsense, non-sceptical, view of knowledge should be left unaffected. Thus a case can be made to accept the implied scepticism (Sturgeon, 1993, 161).

However, there is a pull in the other direction. Quantum indeterminacy equally affects ordinary judgements of possibility. We say that water can pass through a towel, but that a marble cannot. (And we do not mean it in the epistemic sense.) If these judgements express nomic possibility, contemporary physical theory refutes them. But we may instead regard this as a reductio of the idea that the judgements express nomic possibility.

In chapter 7, we defend the following principle:

**Knowledge-possibility link** The sense of possibility relevant to method infallibility is the one that is relevant in ordinary (alethic) possibility judgements, or in some canonical subclass of them.

Knowledge and possibility in the ordinary sense (or in some suitably regimented version of the ordinary sense) go hand in hand. If widespread low chances of bizarre outcomes refute ordinary impossibility claims, they also refute ordinary possibility claims. But if not, they do not. The matter is left open, but there is no straightforward argument that method infallibilism is sceptical.

of qualities. Laws are the statements that satisfy the best combination of simplicity and fit with the distributions; probabilistic laws ensue when perfect fit requires excessively complex laws. Suppose the world consists solely of a series of dice throws. There is no simple pattern in the outcomes, but the overall proportion of each outcome type is about 1/6. This makes true a probabilistic law that assigns each outcome a chance of 1/6. Now the law assigns a non-zero probability to a pure series of 6s. But in the world consisting of the pure series of 6, the laws are different: the pure series in that world makes it true that the law is that a 6 comes out. So the series of 6 is not nomically possible in the first world, even though the laws assign it a non-zero objective chance — call that an impossible future. Such effects are only visible at the global level, however. To falsify the quasi-totality of my beliefs about the future, only small-scale deviations from the actual pattern are needed. Lewis’s impossibility of extremely improbable futures is of no help here.
Ordinary possibilities of error

The hardest cases for the non-sceptical method infallibilist involve cases that would ordinarily be judged to be possible. Vogel has put forward a number of such cases:

**Car Theft** A few hours ago, you parked your car on a side street in a large city. You remember clearly where you left it. Do you know where your car is? We are inclined to say that you do. But hundreds of cars are stolen every day in the major cities of the United States. Do you know that your car has not been stolen and driven away from where you parked it? Many people have the intuition that you would not know that. (Vogel, 1999, 167, see also Vogel, 1990, 15–6)

(Vogel is interested in deductive closure in that case, but I will not discuss this aspect here.) A similar case has been put forward by Sosa:

**Rubbish Chute** On my way to the elevator I release a trash bag down the chute from my high rise condo. Presumably I know my bag will soon be in the basement. But what if, having been released, it still (incredibly) were not to arrive there? That presumably would be because it had been snagged somehow in the chute on the way down (an incredibly rare occurrence), or some such happenstance. But none such could affect my predictive belief as I release it, so I would still predict that the bag would soon arrive in the basement. (Sosa, 2000b, 13)

Let \( \alpha \) be the target case in Vogel’s Car Theft, where the proposition believed is that one’s car is parked on the side street. Suppose that \( \alpha \) is a case of knowledge. There are actual cases \( \beta, \gamma, \ldots \) in which subjects believed that their car was parked on a side street by remembering it,

---

71. See also Vogel’s Rookie cop case (Vogel, 1987, 212, though one may argue that the rookie’s plan of shooting a bullet down the barrel of a gun is impossible in an ordinary sense), his Heart attack, Luncheonette on fire and Meteorite cases (Vogel, 1990, 20–1), and his Night Watchman and Hole-in-one or “Heartbreaker” cases (Vogel, 1999, 164–5). Pappas and Swain’s (1973, 75) Back-up Generator case is an earlier proposal of such a case, but I do not find it very clear or compelling.
but in which they were wrong because their car had been stolen in the meanwhile. Since the cases are actual, they are possible alternatives to α, by the Actual Cases are Possible assumption. Does method infallibility entail that α is not a case of knowledge?

There are two broad ways for the infallibilist to go here. The first is to use proposition centring and/or subject-centring. Proposition-centring: subjects in β, γ believes that their car is parked in such-and-such a street, whereas the subject in α believes that her car is parked in the side street; their car could have been stolen (and in fact was), but her car could not. Subject-centring: cases β and γ are errors of other subjects; what matters is whether the subject in α could be wrong. I do not think the prospects in that direction look good. I have argued that proposition-centring is disastrous (4.1) and I think subject-centring is problematic as well (appendix E).

The alternative is to take the method in α to differ from the one in β, γ, . . . . The difference in methods could be grounded in internal differences — after all, it is unlikely that two actual subjects are ever in the very same internal physical state. But I do not think that the strategy is very promising; the differences are likely to appear insignificant (e.g. in the perceived colour of one’s car). More likely, differences in methods could be grounded in differences in the subject’s environments. The rough idea is that using one’s memory of where one has parked one’s car in a theft-ridden area and period is not the same as using one’s memory in a theft-free area and period. We have sketched a similar approach to the Two Naturalists case (p. 198).

A similar line is open in the Rubbish Chute case (Pritchard, 2005, 164–5). Let α′ be the target case. There are undoubtedly actual cases β′, γ′ and so on in which subjects mistakenly believe that their bag would go down the chute of their building. But if α′ is a case of knowledge, there is a relevant difference between the cases that constitutes a different method.

Can this way out be blocked by making sure the cases involve the same method? One could try to use the “Methods are History” assumption (MH). Consider again the Rubbish Chute case α′. We cannot use as relevant error cases actual cases β′, γ′, . . . , since those do not differ merely
with respect to their future. So we need a case $\alpha^*$ in a non-actual possible world whose history matches actuality up to $t_\alpha$. Moreover, what happens in $\alpha^*$ should be an ordinary possibility for $\alpha$. Thus $\alpha^*$ should be a case in which the very same bag gets stuck in the chute, and it should be true in $\alpha$ that this could happen. I submit that in such a case the intuition that the subject knows disappears.

Vogel and Sosa’s cases give prima facie examples of knowledge despite an ordinary possibility of error on the basis of the same methods. But once we constrain the history to make sure that the error case does not differ in method, it is doubtful that subject knows. It is thus open to argue, contrapositively, that the harmless error possibilities in the original cases did not involve the same methods. Method infallibility is not counter-exampled.

### 4.3.6 Conclusion

We have discussed the two main objections to the necessity of method infallibility for knowledge: the generality problem and its apparent sceptical consequences. The generality problem is the demand for a specification of methods in terms that are independent of knowledge and of applications of the concept of knowledge. The demand cannot be met, but this does not render method infallibilism untestable or contentless. I have argued that we should instead use the method infallibilism hypothesis and our judgements on knowledge as a guide to what methods and infallibility are. Method infallibilism is tested by the success or failure to find natural notions of methods and infallibility on which knowledge requires an infallible method. I have used the strategy to show that methods can be fine-grained, that they need not be luminous, and that they must satisfy a (limited) internalist constraint.

The strategy applies to the sceptical worry as well. Using ordinary judgements about knowledge — and thus in a sense assuming the falsity of scepticism —, we have eliminated one notion of possibility (metaphys-

---

72. More on the relevant notion of possibility in ch. 7.
4.4 Objections to sufficiency

Granting that method infallibility is necessary for knowledge, is it sufficient? Here I review three arguments to think that it is not. The first is that some internal justification is needed as well (4.4.1). The second is that it should also be the case that there are no defeaters for one’s belief (4.4.2). The third is that method infallibility being too narrow or specific, general reliability is needed as well (4.4.3). Though the second and third argument raise difficult issues, they fail to establish that method infallibility is insufficient. As a positive consideration in favour of sufficiency, I offer a maxim of Simplicity according to which one should not easily submit to the temptation of adding conditions to one’s account in order to block counterexamples (4.4.4). The maxim sets the bar higher for stronger requirements on knowledge. They cannot be a conjunction of method infallibility with something else; they have to replace method infallibility altogether.
4.4.1 Internal justification

Bonjour’s (1980) cases against externalism about knowledge can be taken to show that method infallibility is not sufficient.

Clairvoyant Norman, under certain conditions which usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact, the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable. (Bonjour, 1985, 41)

Say that a method of Norman’s belief is using his clairvoyant powers and that those are infallible. If Norman does not know, then method infallibility is not sufficient. 73

I share Vogel’s reaction to the case: “I am not sure exactly what to make of this discussion, since I tend to lose my bearings when clairvoyance is so much as mentioned.” (Vogel, 2000, 608). He suggests (608n11) a useful alternative case to Bonjour’s: perfect pitch. Cleaning up his case a bit, we have: 74

Perfect Pitch Wolfgang has perfect pitch: he can immediately identify what absolute pitch he is hearing. He has never asked himself whether he had such an ability, nor even heard of people having or not having such abilities, and has no reasons for or against the possibility of such a cognitive power. His neighbour regularly plays piano. She changes the pieces she practises every day, and Wolfgang has no reason or evidence for or against the thesis that the first note of her first piece today will be the same as the first note of her

73. See also Lehrer’s (1990, 162–163) Truetemp case and Foley’s chicken sexers (Foley, 1987, 168–9).

74. The cleaning involves: (1) removing the fact that the subject believes herself to have perfect pitch, (2) not requiring that the subject has concepts of tones, which would suggest a musical education.
first piece yesterday. When she starts playing, however, he forms the belief that the first note is the same as the first note yesterday. Wolfgang’s belief results from his perfect pitch and is true.

I do not feel any reluctance to say that Wolfgang knows that the first note is the same.

What are we to make of the contrast between the two cases? First, the Clairvoyant could be taken to show that knowledge requires some higher-order justified belief or knowledge. One has to be precise about the content of the higher-order belief. Knowledge that \( p \) may require that one knows or justifiably believes that:

(4.1) one knows that \( p \).

(4.2) there is an infallible method on which one’s belief that \( p \) is based.

(4.3) \( m \) is an infallible method. (Where \( m \) is the method on which one’s belief is based.)

The Perfect Pitch case shows that (4.3) is not required; Wolfgang does not even have a concept of the perfect pitch method. He has a concept of hearing, but hearing is not infallible. However, the case does not clearly show that (4.1) is not required: it is not clear at all that Wolfgang does not know that he knows that the first note today is the same as the first note yesterday. Now if Wolfgang does a bit of epistemology on the side, and knows that methods infallibilism is true, he may well deduce (4.2) from (4.1) and thereby know or justifiably believe that (4.2). This may entail that Wolfgang could in principle, upon reflection, know or justifiably believe that (4.2). But that falls still short of knowing (4.3), for Wolfgang may not be able to refer to the method he is using. And conversely, the fact that the clairvoyant does not know (4.3) is not sufficient to show that he is not in position to know (4.1) or (4.2). Finally, if it is argued that knowledge requires knowing (4.1) or (4.2), it is not clear that Wolfgang does not have an infallible method for his beliefs that (4.1) and (4.2).

75. If the neighbour comes down and tells Wolfgang that the first note of her first piece today is the same has yesterday, he could truthfully answer: “I knew it”.

76. We present method models that validate the principle that if one knows, one knows that one knows in sec. 5.5.4.
sufficiency is not threatened yet.

What are method infallibilists to say about the original Clairvoyant case? They have to argue that either the Clairvoyant’s methods are fallible or that he knows. Here is one way to do it.

**Clairvoyant pair** In α, Norman forms a belief that the President is in New York, as in BonJour’s original case. He is put to quick sleep and awakes thirty minutes later. He then forms again the belief that the President is in New York, but this time based on a false memory (case β). In β, his belief is false because the President has just left the city.

Now if BonJour thinks that Norman can be in the same internal physical state in β, then we can use the local internalist constraint on methods to say that case β involves the same method. So Norman’s method is not clairvoyance (but rather *forming a spontaneous belief*), so Norman does not know. If his internal state in β has to be substantially different, e.g. if he has a specific visual-like experience in α that he does not have in β, then we say that he knows in α.

Accepting that Wolfgang and Norman can know without independent justification that their belief is based on a reliable source raises issues related with epistemically circular inferences, variously known as epistemic bootstrapping or the problem of easy knowledge. From his knowledge that the two first notes were the same, Wolfgang can deduce that:

(4.4) It is not the case that (the two notes were not the same but they sounded the same to me).

---

77. See p. 196 above for such cases.
4.4. Objections to sufficiency

One should either grant that Wolfgang knows (4.4), reject deductive closure for knowledge, or deny that Wolfgang knows. Contextualists argue that the first option is true of some contexts and the third of others. The problem could ultimately provide support to the idea that method infallibilism is insufficient, if the only viable solution turns out to be to deny that Wolfgang knows. The problem is deep and intriguing, but I will not go into it in the present work. Suffice it to say that it is far from obvious that the only viable solution is the sceptical one.

4.4.2 The Absence of Defeaters

Knowledge may be thought to require the absence of defeaters as well as method infallibility. We say that a proposition is a defeater in a case \( \alpha \) iff the proposition indicates (in a sense to be specified) that one does not know in \( \alpha \). To be a defeater, a proposition has to be either true or believed: we thus have believed defeaters and factual ones. Defenders of the no-defeater condition take the relevant sense of indication to be factive: if there is a defeater in a case \( \alpha \), one does not know in \( \alpha \). Two variants of Williamson’s (2000, 78–9) friendly dog case illustrate the two notions of defeaters:

**Friendly Dog - factual** Charlie sees a dog that he correctly takes to be friendly. However, the dog behaves in ways that would justify the false supposition that it is hostile. So far it has happened to behave in such a way only when Charlie’s back was turned, and Charlie has formed no suspicions yet.

**Friendly Dog - believed** Charlie sees a dog that he correctly takes to be friendly. However, Charlie has been given convincing but false reports that the dog has bitten a number of people. While believing

---

79. See however related discussion about epistemic circularity in Ur-foundationalist and externalist conceptions of knowledge in ch. 2, secs. 2.2.5 and 2.3.2.

80. One could argue that sceptical arguments provide defeaters that defeat by merely being contemplated, without being true nor believed. I leave them aside here.
that the reports are true, Charlie nevertheless trusts his intuition and thinks that the dog is friendly.\footnote{Classical cases of defeaters are Lehrer and Paxson’s Tom Grabbit case (1969, 228) and Harman’s Assassination and Fake Letters cases (1973, 143–4). I rather use Williamson’s for its ordinary setting and the ease with which it can be turned into series.}

The cases can \textit{prima facie} be constructed as cases in which (1) one has formed a belief on the basis of an infallible method, but (2) there is a true or believed proposition that indicates that the belief is false or that the method unreliable, and (3) one does not know. They support the idea that method infallibility is not sufficient. One would need to introduce a no-defeater requirement or an adherence requirement to account for such cases (see 3.4.4 for the latter).

To defend the idea that method infallibility is sufficient, we have to hold that:

\textbf{Real defeaters entail fallibility} Each candidate defeater either shows that the method is fallible or fails to prevent knowledge.

We illustrate with the factual defeater case. Consider a sizable area containing a thousand of dogs. To each dog corresponds a case; Charlie’s case is among them. Suppose that some dogs in the area behave erratically as Charlie’s dog does but are hostile; in the corresponding cases, some subjects are mistaken. If Charlie’s methods are the same as the ones of those subjects, he fails to know. The defeater is a genuine one if it shows such cases to be possible. (If they are, we essentially fall back on a fake-barn style of case.)

By contrast, suppose that there are no similar but hostile dogs in the area. Some dogs behave erratically as Charlie’s dog does, but none of them is hostile. The hostile dogs always behave in a hostile way, for instance. The dog bites stories are unreliable, or not about Charlie’s dog. Then we can argue that Charlie knows, though perhaps he does not know that he knows. The case is analogous to the following:

\textbf{Alleged Hallucination} Stein can clearly see three stains on the table cloth, but his friends argue compellingly that there are only two and
that he is hallucinating. Stein, however, cannot help believing that there are three stains. 82

In such cases I would argue that Stein knows, in spite of his evidence to the contrary. But it may be that he fails to know that he know. This, however, depends on what account is given of higher-order methods. I will not explore the question further here.

4.4.3 Reliability and virtue

We allow methods to be very fine-grained (4.3.4) and we adopt some fairly restricted notion of possibility (4.3.5). The worry arises that we will encounter analogues for methods of the necessary truths problem for belief-centred requirements (4.1.1). For we can build cases in which a subject’s method is infallible in the particular circumstances she is in, and yet she does not know.

Pritchard’s Temperature Room is one such case: 83

Temperature Room Temp forms his beliefs about the temperature in his room by consulting a thermometer on the wall. Unbeknownst to Temp, however, the thermometer is broken and is fluctuating randomly within a given range. Nonetheless, Temp never forms a false belief about the temperature by consulting this thermometer since there is a person hidden in the room, next to the thermostat, whose job it is to ensure that whenever Temp consults the ther-

82. The example is based on Asch’s (1951) famous experiments on conformity. In the experiments, the subject is asked to tell which of three lines is equal in length to a given line. The response is obvious, but the subject answers last after seven fake participants gave their answers. In some cases, all the fake participants give a false answer. (The main difference with our case is that the subject is not given reasons to think that he is mistaken.) On average, one third of subjects are swayed to give the wrong answer, and three quarters followed the group at least once. But few said afterwards that they believed the wrong answers. For those who believed the right answer, it is hard to argue that they did not know the true answer. (At most one could argue that they did not know that they knew.)

83. Luper-Foy’s (1984, 31) Flip case is analogous, except that the subject is irrationally basing his beliefs on coin tosses. See also Hawthorne (2004, 56n).
mometer the temperature in the room corresponds to the reading on the thermometer. (Pritchard, 2010)

Temp does not know. But if we allow one method he is using to be individuated by the kind of environment he is in, that method will be infallible. So method infallibilism appears insufficient. Pritchard (2010) draws a similar conclusion concerning the safety condition, and consequently defends a requirement of virtuous belief alongside safety. 84

The method infallibilist has some elbow room here, however. First, if subject-centring is given up, Temp’s method is fallible if other people could have false beliefs on its basis. If broken thermostats are frequent in the area, this may be the case. Second, ordinary possibilities can be such that it is a possibility, in Temp’s case, that the hidden person is not there. But it is not clear that such moves are available in all cases. Consider:

**RED MUSHROOMS** Bluffon is an amateur naturalist who explores the flora of an island hitherto unknown to him. He spots a few red mushrooms and forms the belief that they are poisonous. The belief is nothing but a hunch based on their unusual colour. As it happens, there are deep evolutionary reasons for which all the red fungi and plants of the island are poisonous.

If we are allowed to individuate a method by the island Bluffon is in (compare p. 198), we may have to say that his belief is based on an infallible method. Yet he cannot come to know, at first sight and on the basis of a hunch, that the mushrooms are poisonous.

Such cases make it very tempting to adopt a double requirement on knowledge such as Goldman’s local and global reliability or a safety and

84. *Revision note.* Duncan Pritchard objected that the discussion in the present section does not go to the heart of the matter, which is that modal conditions cannot capture the fact that knowledge requires a certain direction of fit between belief and fact: belief should be adapted to the fact, and not the opposite.

The matter deserves more discussion, but my short reply is that I do not think that knowledge requires such a direction of fit. It is plausible, or at least conceivable, that in ordinary cases of intentional action, our belief that we are going to do something is partly the cause of our doing it. Yet that does not seem to me to prevent such beliefs to constitute knowledge.
4.4. Objections to sufficiency

virtue view (see 3.7). On such a view, knowledge requires both “local” infallibility but also that the belief be formed on the basis of a virtue or a generally reliable process. Bluffon fails to know by failing to satisfy the second condition: overall, inferring that a plant or fungi is poisonous because it is red is not reliable.

There are strong reasons to resist such temptation, however. I will spell them out and then return to what can be said if we stick to method infallibility alone. We focus on the virtue proposal.

First, the virtue requirement gets us in a generality problem that is much less constrained than the one for methods. Let $\alpha$ be a case of knowledge; then (by hypothesis) it is a case of virtue. Which other cases $\beta$ involve the same virtue? Since virtue in $\alpha$ requires only avoidance of error in most cases $\beta$ in which the same virtue is used, we cannot rule out any case of error, and for many candidate ranges of same-virtue cases, one is hard pressed to say whether error is avoided in most of them, for there is no obvious measure of the proportion, and even if there was, we only have a vague threshold.

Second, the conjunction of the two requirements generates a secondary Gettier problem, that is, cases in which both conditions are satisfied but it is somehow a coincidence that they are. For instance:

**Exceptional red mushrooms** Henri Thorough is one of the brightest naturalists of his time. As he is well aware, there is a certain toxic substance in a number of species of mushrooms that causes them to have a particular shade of red. All known species with that shade of red have the substance in question. Exploring a virgin island, he spots a new species of mushroom that has that particular shade of red and forms the belief that it is poisonous.

The situation with respect to the mushrooms is as before: for deep evolutionary reasons, all the red flora on the island is poisonous.

---

85. Plantinga has defended a similar strategy for his proper function account (Plantinga, 1997, see above sec. 3.2.1).

Revision note. I originally attributed the safety and virtue view to Pritchard, but his recent work clarifies that he does not hold the view. Cf. section 3.7.4.
But the shade of red in that mushroom is due to an entirely different substance from the one Thorough is familiar with.

If Bluffon is a case of infallible method, Thorough is one as well. But in Thorough’s case, the method is in addition generally reliable: there are but a few exceptions to the generalisation on which he relies. I assume that by most accounts of virtue, his inference will also count as virtuous. Yet he does not know. This seems to be because it is a coincidence or accident that both conditions are satisfied, just as in Gettier cases we have a coincidental satisfaction of a justification/reliability condition and a truth condition.

Third, the virtue requirement cannot be simply a matter of reliability. Consider:

**4th Powers** Powell uses the following method to determine whether a number \(x\) is the power to the 4th of a natural number (that is, whether there is a natural number \(y\) such that \(x = y^4\)):

Given any number \(x\), he judges that \(x\) is not the power to the 4th of a natural number.\(^86\)

Given any range of real numbers \((i; j)\) with \(i \neq j\), the method is virtually infallible: 4th powers of natural numbers are natural numbers themselves and the probability of randomly picking out a natural number in \((i; j)\) is 0, so the probability of the method going wrong on a number randomly picked in \((i; j)\) is 0. If we focus on ranges of natural numbers alone, the method is still doing very well as long as some finite range wider than \((0; 3)\) is selected. It is 97% reliable on a natural number randomly picked up between 0 and 100, 99.5% reliable between 0 and 1000, 99.9% between 0 and 10000.

Still the method is not, intuitively, a way in which Powell can come to know that a number is not a 4th power. Not even if it is infallible in the circumstances:

\(^86\) Here are the ten first 4th powers: \(0 = 0^4, 1 = 1^4, 16 = 2^4, 81 = 3^4, 256 = 4^4, 625 = 5^4, 1296 = 6^4, 2401 = 7^4, 4096 = 8^4, 6561 = 9^4\), 10000 = 10^4.
Unbeknownst to Powell, the numbers he is presented with are lower than a billion and cannot include the digits 1 or 6. (Imagine for instance that the numbers are formed from a collection of digits and passed on to Powell, but the collection lacks 1 and 6.) As it turns out, all $4^{th}$ powers below a billion include the digits 1 or 6. So it is impossible for Powell to make a mistake. Yet he does not know that the numbers presented are not $4^{th}$ powers.

The lesson of the case is that sheer reliability is not what makes a belief-forming habit give knowledge. If we call “virtues” belief-forming habits that are apt to give knowledge, virtues are not just reliable faculties.

There are various ways a virtue theorist can deal with the three difficulties. She may reject the generality problem as we did. She can adopt an explanationist account to avoid the secondary Gettier problem (Sosa, 2007, 22–3, Greco, 2003, 116–7, 123): knowledge requires success because of virtue, not just success and virtue. The explanationist account may be sufficient to explain why Powell does not know: one may grant that Powell’s reasoning is virtuous but argue while denying that when he gives a right answer his belief is true because of his reasoning. Alternatively, virtue as distinct from reliability may itself be given an explanationist account, or it could be taken as a primitive. These options deserve further exploration, but the difficulties sketched suggest that a solution in purely method infallibilist terms is well worth a try.

Let us now return to the mushroom cases. The trouble for method infallibilism arises from two assumptions:

(4.5) The naturalists’s methods are to infer, on their island, that a mushroom is poisonous if it is red.

(4.6) The deep evolutionary reasons for which all red flora on the island is poisonous make it impossible that there are non-poisonous red mushrooms on the island.

87. I do not know if the property holds afterwards.

88. Interestingly, statistical methods not unlike Powell’s, but with much lower odds of mistake, are routinely used in applied mathematics (notably cryptology) to establish, for instance, that a particular number is prime.

89. One may think that this is the lottery problem in a new guise.
Given both assumptions, the naturalist’s methods are infallible. To maintain that method infallibility is sufficient for knowledge, one has to reject one or both assumptions.

To reject (4.6), we need to distance the notion of possibility relevant to infallibility from all-purpose notions of nomic, ordinary or close possibilities that are for instance at play in the semantics of counterfactuals (as suggested by Hawthorne, 2004, 56n). For instance, we could force the introduction of normal possibilities which bypass some specificity of the situation. In the case at hand, we would introduce possibilities in which the island contains similar but non-poisonous red mushrooms. I will not explore this option further here.

The alternative is to individuate the naturalists’s methods as being less specific. In particular, we should remember methods are not tied to a particular belief. Thus in Thorough’s case, it is natural to consider that his belief that the mushroom contains the toxic substance he is familiar with is based on the same method as his belief that it is poisonous. But that belief is false, so the method turns out fallible. In Bluffon’s case, it may be that the same method yields also a belief that the mushroom is deadly poisonous, or that it is even poisonous to touch, and so on. This may happen in the actual case or in ordinary counterfactual ones. Some of these beliefs are false, so the method is fallible.

We have discussed cases in which a belief is apparently based on an infallible method, and yet fails to be knowledge. The case suggest that method infallibility is insufficient and that knowledge may additionally require some form of reliability or virtue. But the cases are not conclusive. One can deny that they involve infallible methods.

### 4.4.4 Simplicity

I want to put forward the following as a rule of thumb:

**Simplicity** Conjunctive requirements tend not to be sufficient for knowledge, because they are susceptible of accidental satisfaction.
4.4. Objections to sufficiency

The idea behind Simplicity is the following. Knowledge is a natural condition, in the sense of “natural” in which being a cat is a more natural property than being a cat or a car and than being a cat and a pet of a chief of state. If a requirement on knowledge consists in two conjuncts, there is a risk that the two are satisfied in a case by coincidence. And if the satisfaction of both conditions is a matter of coincidence, the case will probably not be one of knowledge. For knowledge is natural and a coincidence of two conditions is not.

The Gettier problem is one illustration of the rule. The target of the problem is a conjunctive requirement: the requirement that a belief be justified and true. In Gettier cases, the two conjuncts are satisfied by coincidence. Correlatively, the subject does not know.

We have seen a similar problem for the idea that knowledge requires safety/infallibility and virtue. In the Exceptional Red Mushroom case, the subject’s relevant method (as it was initially assumed to be individuated) is both locally infallible and overall virtuous, but it is a coincidence that the two conditions are satisfied.

A similar problem may affect no-defeater conditions. For instance, suppose that it is only required that there is no believed defeater. It may be a pure matter of luck or coincidence (say, a blow on the head) that a person does not believe a defeating proposition in a case. Those who have the intuition that such defeaters prevent knowledge would probably not ascribe knowledge in such a case.

We may call secondary Gettier problems the cases in which a conjunction of requirements is satisfied by coincidence or accident in a way that precludes knowledge. I am not claiming that all conjunctive requirements face secondary Gettier problems. (I have not been able to characterise the problem more formally nor to rigorously identify a subclass of requirements that would face such problems.) But as a rule of thumb, conjunctive requirements tend to face secondary Gettier problems.\(^\text{90}\)

\(^{90}\) Simplicity is related to Williamson’s (2000, 66) notion of primeness, but distinct. Williamson calls composite a condition that is a conjunction of a narrow condition and an environmental condition, and prime a condition that is not composite. A narrow
Secondary Gettier problems warn against a common practice in epistemology. It is usually assumed that when a requirement is necessary but not sufficient, one can keep it as a proper part of a full account. Thus virtually all attempts at analysing knowledge start by listing belief and truth and go on looking for some further requirements. The proper reaction to a counterexample to sufficiency, it is assumed, is to add a further requirement in one’s account. Simplicity suggests exactly the opposite. Each additional requirement is a potential source of secondary Gettier cases, which will show that the conjunction is not sufficient. Adding even further requirements will only lead to further problems. The moral is that one should not add conditions but replace one’s necessary conditions in the face of counterexamples to sufficiency.\textsuperscript{91, 92}

Method infallibilism is a simple condition on knowledge that entails belief and truth:

\begin{itemize}
  \item A narrow condition is such that whenever the subject is in the same total internal physical state in two cases \(\alpha\) and \(\beta\), the condition obtains in either both or none. An environmental condition is such that whenever the external environment is in the same total external physical state in two cases \(\alpha\) and \(\beta\), the condition obtains either in both or none. Given free recombination of narrow and environmental conditions (Williamson, 2000, 73–5), Williamson shows that knowledge cannot be a conjunction of a narrow condition and an environmental one.
  \item Does primeness entail simplicity? No. A conjunction of independent prime conditions is itself prime without being simple. Does simplicity entail primeness? No. Narrow and environmental conditions are trivially composite, since they are equivalent to a conjunction of themselves and the condition that obtains at all cases — which is both narrow and environmental. So simple requirements that state narrow and environmental conditions are not prime. Yet for broad conditions that are neither narrow nor environmental, simplicity plausibly entails primeness. If a broad condition is composite, then it is just the condition of a (non-trivial) narrow condition and a (non-trivial) environmental condition. Given this, we can plausibly find an area of the modal space, so to speak, where they are satisfied by coincidence. If so, composite broad conditions are not simple. Contrapositing: simple requirements that state broad conditions are prime. So the Simplicity rule appears more general than a requirement of primeness. But it is harder to establish. (\textit{Note revised.})
\end{itemize}

\textsuperscript{91} We have seen an illustration of this methodology in section 4.1 when we replaced safe-belief requirements with safe-method ones.

\textsuperscript{92} \textit{Revision note}. Duncan Pritchard objected that adding a virtue condition to a safety condition in the way he did in the account he defends will not lead to secondary Gettier problems. But precisely, his “virtue anti-luck epistemology” does not consist in \textit{conjoining} a virtue condition to a safety condition: the account is that knowledge requires a belief that is \textit{safe because virtuous}, not \textit{safe and virtuous}. The claim here is only that accounts of the latter form are susceptible of coincidental satisfaction.
4.5 Conclusion

**Method infallibilism** $S$ knows that $p$ iff $S$’s belief that $p$ is based on a method that could only yield true beliefs.

Note that, as with parametric requirements in general (3.6), the logical form of the requirement is not a conjunction, but an existential statement:

$$\exists m(Bma \land I_m\alpha),$$

where $I_m\alpha$ abbreviates: $\forall \beta(\alpha \text{R} \beta \to \neg \text{E}_m\beta)$.

Logical form alone does not guarantee that the requirement cannot be satisfied in an accidental manner. But at least the form is not conjunctive like the one of simple modal requirements or double modal requirements.

Given that method infallibility is necessary, Simplicity puts the bar higher for competitor accounts according to which it is not sufficient. Such accounts are ill-advised to settle for some additional requirement that would block apparent counterexamples to the sufficiency of method infallibility. They have to put forward a simple requirement that replaces method infallibility and deals with the additional cases. Of course there is always the option of saying that the only simple replacement is the requirement of knowing itself. But this would acknowledge that method infallibilism is the best non-tautologous requirement.

### 4.5 Conclusion

Modal accounts of knowledge face three problems. The first is the necessary truth problem, the second is the Gettier and lottery problem — which constitute a single problem from the modal point of view —, and the third is the secondary Gettier problem. To avoid the first, modal requirements on knowledge should require avoidance of errors based on a same method. To avoid the second, modal requirements should be infallibilist. To avoid the third, they should only require method infallibility. Together, and on the assumption that a modal account of knowledge is desirable, they suggest that method infallibility is sufficient for knowledge:
Radical Method infallibilism  S knows that p if and only if S’s belief is based on a method that could only yield true beliefs.

(The version is “radical” because it states that methods infallibility is both necessary and sufficient for knowledge. The “moderate” one states only that it is necessary.)

Method infallibilism raises two worries. First, it is unclear how methods are individuated. This leads to the Generality problem. Second, it appears to have implausible sceptical consequences. Our answer to both problems relies on the idea that the relevant notions of methods and infallibility should not be taken to be given independently of our judgements about knowledge and the role they play in the method infallibilist account of knowledge. Rather, they are partly picked out by the method infallibility requirement itself: methods are any properties of belief cases such that their infallibility entails knowledge. We have argued that this methodology does not prevent method infallibilism from being a testable and reductive account of knowledge. We have applied the methodology to formulate some constraints on the individuation of methods and on the relevant notion of possibility.
Part III

Formal Epistemology
Chapter 5

Methods-based Models

In this chapter we introduce a new formal representation of knowledge and belief based on method infallibilism. We introduce formal models of methods and their infallibility, and then define belief and knowledge in terms of them. The resulting models are a significant extension of so-called “neighbourhood models”. We argue that epistemological notions and problems like Gettier, inductive knowledge, fallible justification, epistemic contextualism, or failure of logical omniscience are represented in a more satisfactory ways in these models than in standard epistemic logic. In general, our models only validate the claim that knowledge is true belief; but we show that a full S5 system can be derived from a set of natural idealisations. The derivation provides some explanation of why and when the S5 axioms should hold, and a vindication of their use. We additionally derive various intuitive connexions between belief and knowledge, as well as some properties of knowledge that can only be stated in reference to methods, such as deductive closure (as opposed to logical omniscience).

The chapter can be read independently of the previous ones, though chapters 3 and 4 provide the philosophical background of the model. The introduction (section 5.1) summarizes the philosophical motivation, however, with an eye to the unsatisfactory aspects of standard epistemic logic. Section 5.2 introduces the formal representations of methods and infalli-
bility. We characterise them in the most general way, without assuming a specific notion of proposition. We define operations on methods and give an algebra for them. Section 5.3 applies the method-infallibilist approach to formalise Gettier cases, fallible justification and inductive knowledge. Section 5.4 introduces methods models properly. For concreteness and simplicity we take propositions to be sets of possible worlds. That creates trouble with so-called Frege cases, though not as straightforwardly as expected. The resulting models are an extension of neighbourhood models, but more explanatory than the latter. (A detailed comparison is made in appendix C.2). We introduce a language for methods-based belief and knowledge. Section 5.5 details the main consequences of the models. In general, the models only validate the uncontroversial claim that knowledge is true belief; however, we derive a full S5 system for a series of natural idealisations of the agent’s methods. The derivation provides a illuminating perspective on why and when the axioms of standard epistemic logic hold.

5.1 Introduction

My starting point is the following puzzle. (1) Whether one knows centrally depends on what basis one’s belief has. (2) Standard epistemic logic cannot represent bases of belief. (3) Standard epistemic logic adequately models knowledge in a number of applications. I introduce formal models of knowledge directly stemming from the idea that knowledge is a matter of bases of belief, or as I will call them, methods. The models are an extension of Scott’s (1970) and Montague’s (1968; 1970) neighbourhood models, and they differ from other recent formal systems aimed at dealing with similar issues. They solve the puzzle by providing an insight into why and when axioms of standard epistemic logic hold, including the most controversial ones. But most importantly, they provide a philo-

sophically satisfying representation of knowledge. We illustrate the point by using them to formalise several classical epistemological problems and views: the Gettier problem, inductive knowledge, deductive closure without logical omniscience, Frege cases, epistemic contextualism, and failure of knowledge to iterate.

Let me first illustrate the puzzle. Consider three cases:

1. **Tea leaves.** As things are, reading tea leaves is not a reliable way to find the truth. My uncle believes on the sole basis of tea leaves readings that I will get a pay raise soon. Whether or not I will, he does not know that I will.

2. **Watson.** Holmes and Watson know the same facts about a case. Reasoning carefully, Holmes deduces that the father is the culprit. Watson is also convinced of this, but on the sole basis of the father’s shady looks. Watson does not know that the father did it.

3. **Induction.** Seeing that the light is on at the neighbour’s, my mother infers that the neighbours are home. In suitably normal circumstances, she comes to know that the neighbours are home.

Here are a few *prima facie* intuitive things to say about the cases. In (1)-(2), my uncle and Watson fail to know because the bases of their beliefs are not adequate for knowledge. In Watson’s case, that is so even though his belief is true and he knows facts which together entail it. In case (3), my mother comes to know something on a basis which fails to entail it. That is so because her basis is adequate or sufficient given the circumstances she is in. All three cases show that consideration of the basis of one’s belief and its adequateness to the circumstances are central to whether one knows.

Given the apparently central role of bases of belief in epistemology, it is striking how hard it is to accommodate the notion in standard models for knowledge introduced by Hintikka (1962). Such models essentially characterise knowledge and belief in terms of *elimination of possibilities*:

What the concept of knowledge involves in a purely logical perspective is thus a dichotomy of the space of all possible
scenarios into those that are compatible with what I know and those that are incompatible with my knowledge. (Hintikka, 2007, 15)

As Hintikka makes clear, the account is not reductive, since the “elimination” of possibilities is itself defined in terms of knowledge. In fact, standard epistemic logic is best construed as a representation of the content of one’s knowledge rather than as a representation of the state of knowing. That does not mean that it does not say anything about knowledge. The problem is rather that what it says seems false, namely that one knows $p$ iff one knows something incompatible with $\neg p$. In our Watson case, Watson fails to know that the father is the culprit even though he knows things that are incompatible with it being false. In our Induction case, my mother appears to learn that the neighbours are home on the basis of facts compatible with them not being here. (One may of course reply that before she came to learn it, the facts known to her where indeed compatible with them not being there, but then she did not know they were there; while since she has drawn the inference, she knows some fact incompatible with them not being there: simply, that they are there. Be that as it may, we still lack an account of how an inductive inference can achieve such a result.) As we pointed out, a natural first step in thinking about these cases is to formulate them in terms of bases of belief or methods. But it is unclear how to introduce such a notion in standard models. In my view, that goes a long way towards explaining the widely

---

2. Lewis (1996) attempts to turn the idea into a reductive account by construing elimination as metaphysical incompatibility with one’s being in the total experiential state on is in. The move has unpalatable consequences (see e.g. Hawthorne, 2004, 60n).

3. See Hintikka (2007, 16): “Epistemic logic presupposes essentially only the dichotomy between epistemically possible and epistemically excluded scenarios. How this dichotomy is drawn is a question pertaining to the definition of knowledge. However, we do not need to know this definition in doing epistemic logic. Thus the logic and the semantics of knowledge can be understood independently of any explicit definition of knowledge. Hence it should not be surprising to see that a similar semantics and a similar logic can be developed for other epistemic notions—for instance, belief, information, memory, and even perception. This is an instance of a general law holding for propositional attitudes. This law says that the content of a propositional attitude can be specified independently of differences between different attitudes.”
noted gap between epistemic logic and epistemology (Hendricks, 2006; van Benthem, 2006): epistemologists mainly think of knowledge as adequately based belief, while epistemic logic represents it as the elimination of possibilities, and it is unclear how to fit the two pictures together.

I am not claiming that epistemic logic cannot be amended and fruitfully extended to deal with some methods-related aspects of knowledge. Much has already be done in that respect. But I take a different approach here. Instead of adapting standard models to specific philosophical purposes, I start from a philosophical characterisation of knowledge and build a model suited to it.

Call *methods* whatever bases of beliefs or ways of forming or sustaining beliefs are relevant to knowledge attribution. Methods need not be conscious procedures one follows step by step. They may (and typically will) essentially involve unconscious computational processes. They may (and typically will) be individuated in a quite fine-grained way, such as the method by which one judges by sight whether a stone is a diamond at a short distance under a good light. They may or may not be individuated “externally” or “broadly”, that is, they may or may not be constituted by facts involving one’s environment. We need not settle these important issues here. As will become clear, a lot can be said about methods while maintaining a fairly abstract perspective on them. Ultimately, the hope is to latch onto methods by building a concept that characterises them structurally and in terms of their relation to belief and knowledge, rather than by attempting to provide a reductive analysis in psychological terms, for instance. But for present purposes, it is sufficient that our model is simply compatible with various substantive conceptions of what knowledge-relevant methods are.

(One important presupposition about methods, though, is that we have equivalence classes of beliefs with the same basis, and not merely a non-transitive similarity relation over particular bases. We leave for

---

4. See e.g. Fagin et al. (1995, chap. 9,10).
5. That use of the term has some currency in epistemology since Nozick (1981, chap.3).
further work the exploration of how our models may be recast in the less demanding setting of a similarity relation between token bases.)

Our guiding idea is that knowledge is belief based on an infallible method:

**Methods infallibilism** An agent knows that \( p \) iff she believes that \( p \) on the basis of a method that could only yield true beliefs.

We say that a method “yields” a belief whenever it forms it or sustains it. A method that produces truth-valueless beliefs (if there are such) counts as fallible by the definition. The “could” is meant to be understood as a (possibly context-sensitive) alethic modal.

There are several reasons to adopt the method-infallibilist account, which we can only mention here. First, it automatically secures factivity and the idea that deduction preserves knowledge – which is to be distinguished from logical omniscience, as we will see. Second, analyses that allow an adequate basis for belief to be compatible with error at the circumstances at hand seem to systematically face Gettier-style counterexamples; some infallibility condition appears required to avoid them (Sturgeon, 1993). Third, it provides a simple diagnosis of why no matter how high the odds, we fail to know in advance that a ticket in a fair lottery is a loser (Hawthorne, 2004). I also hope that the intuitive results we get from the formal implementation of the account will further support it. But be that as it may, the models we introduce can alternatively be extended to represent fallibilist notions of knowledge.

A crucial aspect of the methods approach is that in order to evaluate whether a particular belief is knowledge, we have to consider other beliefs one has or could have had on the same basis. Consider in particular:

1. **Fake oranges.** Looking at a particular orange in a fruit bowl, Oscar believes that *that orange* is a fruit. Unbeknownst to him, the other “fruits” in the bowl are perfect wax replicas of oranges.

2. **Prime numbers.** Primo has a mistaken way of evaluating whether a number higher than 20 is prime: he adds its digits, and if the sum

---

is prime he judges that the original number was too. For instance, since $4 + 7 = 11$ is prime, he (rightly) believes that 47 is prime.

Assuming essentialism, there is no possible world where that orange, the one Oscar is looking at, is not a fruit. And there is no possible world where 47 is not prime or where Primo’s method would lead him to wrongly believe that it is not. Yet both fail to know, because they could have believed false propositions on the basis of the same methods: Oscar would falsely believe on the same basis that an “orange” next to the one he is looking at was a fruit too, and Primo would falsely believe on the same basis that 49 is prime. The fact that, so to speak, we evaluate a belief by looking at whether it is in “good” or “bad” company is the basic idea of our formalisation of methods.

5.2 Methods

5.2.1 The space of possible methods

A method is something that yields beliefs. But it does not need to yield the same beliefs wherever it exists; in fact, it typically yield beliefs as a function of other factors. First, it may yield beliefs as a function of the state of the world. Facing a table, Alice opens her eyes. She immediately forms beliefs: say, that there is an apple, or that there is an apple and that there is a pear, depending on what there is on the table. That is the idea of a purely non-inferential method, and it is naturally modelled as a function from worlds to sets of propositions. Second, it may yield beliefs as a function of other beliefs. For instance, modus ponens would lead you to believe that $q$ from the premises that $p$ and that $p \rightarrow q$. That is the idea of a purely inferential method, and it is naturally modelled as a function from sets of propositions (premises) to sets of propositions (conclusions).

Generalising both ideas, a purely non-inferential method is a function from worlds and sets of premises that is constant over sets of premises. It yields a set of “conclusions” depending on the state of the world, for any premises whatsoever, including no premise at all. A purely inferential
method is a function from worlds and sets of premises that is constant over worlds. It yields the same conclusions for a given set of premises, whatever world one is in. Mixed methods (if there are any) are variable functions from worlds and sets of premises to sets of conclusions.\footnote{Here are two candidates cases of mixed methods: (1) perceptual processes that are sensitive to one’s background beliefs; (2) trust processes (roughly, inferences from \textit{S said} \textit{p} to \textit{p}) that are sensitive to subtle visual clues in a quasi-perceptual way. But whether or not one wants to contend such methods does not matter for the purposes of this paper.}

We are now in position to define the space of all possible methods as the space of all functions from worlds to functions from sets of premises to sets of conclusions. Let $W$ be a set of worlds and $P$ a set of propositions:

**Definition 5.1.** $M = W \rightarrow (P(P) \rightarrow P(P))$ is the space of possible methods.\footnote{An alternative notation is $W^{P(P)^P}$.}

Each method $m \in M$ maps each world $w \in W$ to a function $m(w)$, which in turns maps each set of propositions $\pi \in P(P)$ to a set of propositions $\pi' \in P(P)$.

**Terminology.** When $\pi' = m(w)(\pi)$, we say that $\pi'$ is the set of conclusions reached by method $m$ at world $w$ from the set of premises $\pi$,
when $p \in m(w)(\pi)$, we say that $p$ is a conclusion reached by $m$ at $w$ from $\pi$,
when $p \in m(w)(\emptyset)$, we say that $p$ is an unconditional output of $m$ at $w$, that is, a conclusion reached by $m$ at $w$ without premises.

**Notation.** We abbreviate:
\begin{align*}
m(w, \pi) & := m(w)(\pi), \\
m(w) & := m(w)(\emptyset).
\end{align*}

**Remark 5.1.** By convention, worlds are noted $w, w', \ldots$, propositions $p, p', \ldots, q, q', \ldots$, sets of premises $\pi, \pi', \ldots$, and methods $m, m', \ldots, n, n', \ldots$. We typically omit domains when they are clearly indicated by this convention. Thus we write: “for all $w$”, “$\forall w(...)$” and “$\{w : \ldots\}$” instead of “for all $w$ in $W$”, “$\forall w \in W$” and “$\{w \in W : \ldots\}$”, and similarly for any $p \in P, \pi \subseteq P$ and $m \in M$.

Now the idea that methods are functions from worlds is a gross simplification. Suppose two agents are in distinct rooms, each facing a table: Alice sees an apple, Bob a pear. Intuitively, we would like to say that the
very same method can produce a belief that there is an apple in Alice’s mind, and no such belief in Bob’s mind. Also, it is natural to take the outputs of a method to depend on the time and not just the world. To model these phenomena, methods should rather be functions from centered worlds $< c, w >$ where $w$ is a world and $c$ a perspective on that world. A perspective is a point from which methods can be applied; a time and a place, at least. Perspectives need not only be where some agent is: a visual method can be fallible in virtue of producing false beliefs as used from the top of the mountain, even though nobody has been, is or will be at the top of the mountain. The life of an agent is then the series of perspectives the agent occupies. The agent’s belief and knowledge are then derived from its life and the methods she has.

These refinements for perspectives and lives are relatively straightforward to introduce, but they needlessly complicate the models for our present purposes. We thus stick to the characterisation of methods in terms of functions from worlds.

### 5.2.2 Operations on methods

For any methods $m, n$, we define:

**Definition 5.2.** The union of $m$ and $n$ is the method $(m + n)$ given by $(m + n)(w, \pi) := m(w, \pi) \cup n(w, \pi)$ for any $w, \pi$.

The composition of $m$ and $n$ is the method $(m \circ n)$ given by $(m \circ n)(w, \pi) := m(w, n(w, \pi))$ for any $w, \pi$.

Method union is the idea that an agent is able to pool together the outputs of different methods. If, given premises $\pi$, $m$ outputs $\{p\}$ and $n$ outputs $\{q\}$, the method $(m+n)$ outputs $\{p, q\}$. Putting limitations on union is thus a way to model the modularity of an agent. For instance, a limited number of unions corresponds to an agent with limited working memory, who is unable to put to use all her beliefs at once. And a systematic bar on uniting certain methods corresponds to an agent whose bodies of beliefs are partly isolated from each other.
Method composition is the idea that an agent is able to apply one method to the output of another: if \( n \) outputs \( \{p\} \) from no premises, and \( m \) outputs \( \{q\} \) from \( \{p\} \), then \( m \circ n \) outputs \( \{q\} \) from no premises. Limits on an agent’s ability to compose methods is thus a way to represent \textit{computationally bounded} agents, such as agents who are only able to go through proofs with a limited number of steps.

Given a set of methods \( M \), we can define its union and composition closure as the smallest set \( M^+ \) such that \( M \subseteq M^+ \) and for any \( m, n \in M^+ \), \( (m + n), (m \circ n) \in M^+ \). \( M^+ \) is the set of methods available to an agent that has the \( M \) methods and is neither modular nor computationally bounded.\(^9\)

Method union and composition are interpreted in a synchronic way here. If a (non-modular) agent has \( m \) and \( n \) at a given time, then \( (m + n) \) is available to her at the very same time. However, union and composition could easily be construed as dynamic processes. For instance, given a certain method set \( M_0 \), one could consider the series \( M_1, M_2, \ldots \) where each \( M_k \) corresponds to applying one step of union and composition to \( M_{k-1}: M_k = M_{k-1} \cup \{(m + n) : m, n \in M_{k-1}\} \cup \{(m \circ n) : m, n \in M_{k-1}\} \) for \( k \geq 1 \). But we will not get into such models here.

Two methods are distinguished by union and composition. The empty method is the identity element for union: \( 0 + m = m + 0 = m \) for any \( m \). The identity method is the identity element for composition: \( 1 \circ m = m \circ 1 = m \) for any \( m \).

\( \langle M, +, \circ \rangle \) is an algebraic structure over the set of methods. We detail its main properties in Appendix C.1. The most remarkable are:

\(^9\) \textit{Revision note.} As defined here, the union and composition closure of a set \( M \) need not contain the identity methods for union and combination defined in our algebra below. The definition could be amended so that these methods (called 0 and 1, respectively). None of our results would be affected. In any case where the agent’s basic method set \( M^\emptyset \) contains some method other than 0 and 1, the closure of \( M^\emptyset \) is identical to the closure of \( M^\emptyset \cup \{0, 1\} \). In the special cases in which \( M^\emptyset \) is empty, or contains only 0 or 1, there is a difference in the result closure set: \( \emptyset, \{0\} \) or \( \{1\} \) respectively in the present definition, \( \{0, 1\} \) in the amended one. But in both cases the resulting agent does not have (unconditional) beliefs at any world, so the difference does not affect results about belief and knowledge.
5.2. Methods

1. Method union is associative, commutative and idempotent:
   \[(m + n) + r = m + (n + r)\].
   \[m + n = n + m\].
   \[m + m = m\].

2. Method combination is associative, but not commutative nor idempotent.
   \[m \circ (n \circ r) = (m \circ n) \circ r\].

3. Combination distributes right-to-left over union, but not left-to-right:
   \[(m + n) \circ r = (m \circ r) + (n \circ r)\],
   but \[m \circ (n + r) = (m \circ n) + (m \circ r)\] may fail.

The fact that combination does not distribute left to right over union is a reflection of the fact that method combination keeps track of information processing or, more accurately, of information dependencies. \(m \circ (n + r)\) corresponds to uniting \(n\) and \(r\) and then applying \(m\), which is not the same as applying \(m\) to \(n\) and to \(r\) separately. To take a simple example: suppose \(m\) outputs the conjunction of any premises and suppose \(n\) outputs only \(p\) and \(r\) outputs only \(q\). The union of \(n\) and \(r\) outputs both \(p\) and \(q\), so \(m \circ (n + r)\) will output \(p \land q\). But applying \(m\) to the premise \(p\) and separately to the premise \(q\) will not yield the conjunction of the two, so \((m \circ n) + (m \circ r)\) will not output \(p \land q\).

5.2.3 Infallibility

A method is infallible if it could only yield true beliefs. Only unconditional outputs of a method matter for infallibility. The fact that \textit{modus ponens} sometimes produces false conclusions given false premises does not show that \textit{modus ponens} is fallible.\(^\text{10}\)

\(^{10}\) Revision note. Timothy Williamson suggested that infallibility should take into account the inferential component of a method as well, namely: a method is infallible only if it would not reach false conclusions from true premises. From this amendment it follows that the union and combination of infallible methods are infallible. I have included this amendment in revised versions of the chapter submitted for publication. However, the amendment is not compatible with the way psychological introspection
The relevant modality is alethic, not epistemic or doxastic. For a method to be infallible, it is not required that one knows or believes that it is. It is sufficient that error is in fact impossible. Within the domain of alethic modalities, many options are open. For instance, one could require that error be physically impossible given the makeup of the agent and the situation she is in (Armstrong, 1973, 168), or that error be impossible in sufficiently similar cases (Williamson, 2000, 100), or that error be impossible in a contextually determined set of relevant possibilities (Lewis, 1996, 553–4). We need not decide between them here. Three points should be mentioned, though.

Infallibility requires at least no error at the actual world: a method that actually yields a false belief is not such that it could not yield one. That is a consequence of the fact that the relevant modality is alethic.

Infallibility need not require impossibility of error across all worlds. In some worlds, pigs can fly, in others, they cannot. Similarly, some methods may be fallible at some worlds, and yet infallible at others.\footnote{In applications of epistemic logic, the set of possible worlds is in effect restricted to a relevant set of worlds. The possibility that the prisoners communicate is ignored, for instance. In such a set up, infallibility across “all” worlds is in effect a restricted form of impossibility of errors. In other applications, the fact that a method can be infallible at a world but fallible at another will play a role.}

An important feature of the notion of possibility is whether what is possibly possible is possible (axiom 4 of modal logic). Physical possibilities and related notions have this property. But if possibility is a matter of sufficient similarity between cases, it does not: a possibility may be sufficiently similar to a second which is sufficiently similar to a third without the first being similar to the third. When the property holds, infallible methods are necessarily infallible; when it does not, infallible methods may be possibly fallible. As in standard epistemic logic, the property turns out to be crucial to derive knowledge of one’s knowledge.\footnote{See Williamson (2000, ch.5) on the failures of epistemic introspection that result when the relevant notion of impossibility does not iterate.}

We model the notion in the standard way by an accessibility relation
over worlds. A method is infallible at a world if its unconditional outputs at all accessible worlds are true:

**Definition 5.3.** Let $W$ be the set of worlds, and $R \subseteq W \times W$ a reflexive accessibility relation over worlds. A method $m$ is infallible at a world $w$ iff:

For any $w', p$, if $wRw'$ and $p \in m(w')$, $p$ is true at $w'$.\(^{13}\)

Different constraints on the $R$ relation give different notions of infallibility.\(^{14}\)

1. $wRw'$ iff $w' = w$: a *true-belief-like* notion of knowledge, in which only the actual world needs to be considered. Though implausible as a philosophical account of knowledge, it is noteworthy that all our results can be obtained in that simple setting.\(^{15}\)

2. $wRw'$ for any $w, w'$: a notion of knowledge that requires the metaphysical impossibility of error, and correspondingly precludes inductive knowledge. Arguably the notion defended by Descartes.

3. $wRw'$ iff $w'$ is “close” to $w$: a *safety* notion of knowledge such as the one defended by Williamson (2000) and Sosa (1996). If closeness is not transitive, what is possibly possible need not be possible, and knowledge of one’s knowledge is not guaranteed (section 5.15).\(^{16}\)

4. In each context $c$, “knows” is associated with a specific $R_c$ relation: a contextualist account along the lines of the one defended by DeRose (1995) and Lewis (1996).\(^{17}\)

---

\(^{13}\) Recall that $m(w') = m(w', \emptyset)$ is the unconditional output of $m$ at $w'$.

\(^{14}\) See ch. 3 for a more detailed study of the various proposals in the literature.

\(^{15}\) Note in particular that that option does not *equate* true belief and knowledge. Suppose that a method outputs both $p$ and $q$, and that $p$ is true but $q$ false. Then one’s belief that $p$ on that basis is not knowledge, even though $p$ is true. The option makes knowledge “true-belief-like”, though, because it would classify many lucky true beliefs as “knowledge” just because no agent happens to use a given method in unfavourable circumstances.

\(^{16}\) See sec. 3.4.5 and 3.4.6. Williamson’s similar-case safety requires a similarity relation across cases. The cases can be represented as worlds in methods models, so that the accessibility relation plays the role of the relation of similarity across cases. But that is a fix, not a proper representation. We leave the latter for further investigation.

\(^{17}\) See sec. 3.4.3 and 6.6.2.
Our results are independent of the choice. They only require that $R$ be reflexive, except knowledge of one’s knowledge which additionally requires transitivity.\footnote{Some related accounts of knowledge cannot be represented without substantial modifications of our apparatus. On Nozick’s (1981, chap.3) view, one knows that $p$ only if: if $p$ had been false, one would not have believed $p$. On the Lewis-Stalnaker semantics of counterfactuals (Stalnaker, 1968; Lewis, 1973), the conditional is true if the corresponding $p \rightarrow q$ holds at each world up to the “closest” $p$ world(s). This means that the range of possibility one has to look at for a given knowledge ascription depends on the particular proposition at stake, in our case, $p$. To model this, one needs to relativise accessibility to propositions: $wR_p w'$ iff $w'$ is at least as close to $w$ as the first $p$ world that is closest to $w$. Infallibility needs to be redefined: $m$ is infallible at $w$ with respect to $p$ iff for any $w', q$, if $wR_p w'$ and $q \in m(w')$ then $q$ is true at $w'$. This invalidates our proof (below) that if a method is infallible at $w$, then the combination of Deduction and that method is infallible at $w$, since it may happen that a method is infallible with respect to $p$, without being infallible with respect to a proposition $q$ deduced from $p$, if the first $q$-world is further away than the first $p$-world. That is why Deductive closure fails in Nozick’s system. Note than on the von Fintel-Gillies semantics of counterfactuals (von Fintel, 2001; Gillies, 2007), the condition is true iff $p \rightarrow q$ holds throughout a set of close worlds fixed by context. The set does not depend on the particular $p$ evaluated. In that setting Nozick’s condition is modelled in our system by a context-relative $R_c$, and it does not violate closure.
}

\section*{5.3 Applications}

We can already sketch how to use methods to model a few epistemological ideas.
5.3. Applications

5.3.1 The prime number case and fake-barn-style Gettier cases

In our Prime Numbers case (p. 230), Primo uses a method \( m \) that both produces a true belief that 47 is prime and a false belief that 49 is prime. (Let us assume that he considered both questions at the actual world.) Let \( P(47) \) and \( P(49) \) be the relevant propositions:

\[
\begin{align*}
\text{\( P(49) \)} & \quad \text{\( P(47) \)} \\
\text{\( w_1 \)} & \quad \text{\( m \)} & \quad \text{\( m \)} & \quad \text{\( \overline{w_1} \)}
\end{align*}
\]

Figure 5.1: The prime number case

(Illustrations for methods models. To avoid cluttering, I lay out the set of possible worlds \( W \) in a column that is repeated horizontally as many times as needed. Propositions are represented by circled sets of worlds; typically we use one column per proposition. Only the unconditional outputs of methods are represented: an arrow named \( m \) between \( w \) and \( p \) indicates that \( p \in m(w) \), that is, \( p \in m(w, \emptyset) \). When the output is true, the arrow is horizontal. Diagonal arrows indicate false beliefs and are signalled by wavy lines. When the output proposition is the empty set, as \( P(49) \) is here, the arrow points directly to its name instead of a circle. When needed, accessibility relations between worlds are represented by dotted arrows labeled with \( R \).)

At \( w_1 \), \( m \) outputs the belief \( P(47) \) that is true at \( w_1 \), but it also outputs the belief \( P(49) \), which is false. Consequently, \( m \) is fallible, and no belief based on \( m \) can be knowledge.

Similar models can be given for the fake-barn style of case (p. 230). Suppose that at \( w_2 \) the agent looks at the real orange, but that there is an accessible world \( w_1 \) where she looks at a fake one and forms the false belief that it is an orange. Write \( O(o) \) and \( O(f) \) for the relevant propositions. (See figure 5.2). \( m \) produces only a true belief at \( w_2 \), but there is an accessible world in which it produces a false belief, namely \( w_1 \), and that is why the
subject fails to know even in world \( w_2 \).

\[
\begin{array}{cc}
O(f) & O(o) \\
\downarrow & \downarrow \\
\bar{w}_1 & \bar{w}_1 \\
\end{array}
\]

Figure 5.2: The fake orange case

The models are straightforward but not trivial. In standard epistemic logic, \( Kp \) holds iff \( p \) is true at all accessible worlds. Unless impossible worlds are introduced, it is true at every world that 47 is prime. So however accessibility is fixed, we get the result that it is known. Similarly, in Fitting’s models for Logic of Proofs (Fitting, 2005, 4), \( t : p \) (which we can read as “\( t \) is the subject’s justification for \( p \)”) holds at a world iff \( t \) is evidence for \( p \) and \( p \) holds at all accessible worlds.\(^{19}\) Here we get the result that it is know that 47 is prime as soon as some justification supports that belief, since the proposition that 47 is prime holds at every world. If we count Primo’s calculations as justifications, we get the wrong results; if we don’t, the models do not explain why they do not count as justifications.\(^{20}\)

19. The same holds for the extension of Fitting’s models presented by Artemov and Nogina (2005, 1066).

20. On Fitting’s (Fitting, 2005, 5) strong models, any formula that is true at all accessible worlds has a justification, so Primo’s belief would come out as justified. Artemov (2008, section 6) suggests that a knowledge-level justification for \( p \) is a “factive justification”, i.e. “sufficient for an agent to conclude that \( p \) is true”. This can be understood in three ways. (a) sufficient for an agent to know that \( p \) is true: the characterisation is circular. (b) such that necessarily, if \( p \) is justified by that justification, \( p \) is true: this wrongly ascribes knowledge in the Prime Number case. (c) such that for any proposition \( p' \), necessarily, if \( p' \) is justified by that justification, \( p' \) is true: this is infallibilism as we have defined it. Artemov’s semantics (based on Fitting, 2005) suggests that he adopts the second construal.
5.3.2 Standard Gettier cases

Consider (a slight variant of) Chisholm’s (1966, 23n) sheep case: a man comes to believe that there is a sheep in a field by seeing a sheep-looking rock in the distance. As it happens, there is one, but it is hidden behind the rock. The case is straightforwardly modelled if we assume that the subject could have formed the same belief on the same basis in the absence of sheep. Let \( s \) be the proposition that there is a sheep in the field. (See figure 5.3.)

\[
\begin{align*}
\text{At } w_2, & \text{ the subject’s method produces a false belief. At } w_1, \text{ } w_2 \text{ is an accessible possibility, so his method is fallible and he fails to know.} \\
\text{The original Gettier (1963) cases are modelled in basically the same way, except that we need to introduce a inferential step. Smith has good evidence that Jones owns a Ford, and infers that Jones owns a Ford (j) or Brown is in Barcelona (b). As it happens, b is true but j is false. Let } \text{m be the method that leads Smith to form the belief that Jones owns a Ford, and let } \text{n be the method that leads him to infer } j \lor b \text{ from } j. \text{ Assuming that the subject could have formed the same beliefs while } b \text{ was false, we have the model illustrated by figure 5.4.} \\
\text{At each world, by method } m, \text{ the subject forms the belief that } j, \text{ and by the method } n \text{ applied to } m, \text{ she infers } j \lor b. \text{ In } w_1 \text{ her evidence is not misleading: } j \text{ is true. In } w_2 \text{ her evidence is misleading, however } j \lor b \text{ is true because } b \text{ is true. That is the Gettier situation. In } w_3 \text{ the evidence is misleading as in } w_2, \text{ but now } b \text{ is not true, so } n \circ m \text{ outputs a false belief. Since } w_3 \text{ is accessible from } w_2, \text{ } n \circ m \text{ is not infallible at } w_2, \text{ and that is why the subject fails to know.}
\end{align*}
\]
Again the result is straightforward but not trivial. A common diagnosis of Gettier’s original cases is the “no-false-lemma” view according to which reasoning from false premises cannot provide knowledge (Clark, 1963). Our models assume nothing of the kind: what matters is whether the subject’s total inference \( (n \circ m) \) could have lead to a false belief, not whether some intermediate steps are. Suppose I slightly overestimate heights, and from my belief that the door is over 2.5 meters high, I infer cautiously that it is at least more than 2m high: we may grant knowledge of the latter even though the door was 2.4 meter high, for instance, and my initial belief false. (See Unger, 1968, 165 for a similar view.) Such verdicts are available in methods models.

### 5.3.3 Fallible but reliable methods

We have modelled Gettier cases in terms of fallible methods. That explains why they are not cases of knowledge, but not why they are not simply cases of unjustified belief, such as belief based on tea leaves.

---

21. Artemov’s formalisation of the cases in the context of the Logic of Proofs endorses a strong version of the no-false-lemma view: namely, that in the sense of justification relevant to knowledge there is no justification of a false proposition (Artemov, 2008, section 6). That threatens the possibility of inductive knowledge (see section 5.3.4). See Lycan (2006, section 6.1) for a recent defence of a generalisation of the no-false-lemma view.
5.3. Applications

We will not implement it formally here, but there is a natural way to introduce the idea: each method gets assigned a reliability measure (a real between 0 and 1) at a world depending on its tendency to produce true beliefs rather than non-true ones at accessible worlds. In a finite setting, we could for instance take the measure to be the ratio of true beliefs to all beliefs produced at accessible worlds, but in general we need not assume that reliability is reducible to other notions except in very simple cases. For instance, if \( m \) is the method that leads one to believe that each ticket in a one-hundred ticket lottery is a loser, we may say that the reliability of \( m \) is .99. This gives a sense in which one’s belief based on \( m \) that one’s ticket will lose is justified without being knowledge.

In section 5.5.1, we discuss how the notion of fallible justification, when plugged in a justified-true-belief account of knowledge, results in Gettier-type cases.

5.3.4 Inductive knowledge

Things being as they are, my mother often comes to know that the neighbours are home by seeing that their light is on. Absolutely speaking, it is of course physically possible that the neighbour’s light is on and they are not there. But given their habits and the general circumstances, they could not be out with the light left on. This is what allows my mother to come to know that they are home simply by seeing their light. In the method infallibilist setting, that is cashed out as the idea that inductive knowledge is a matter of local infallibility, that is infallibility over a relevantly restricted set of accessible worlds.\(^{22}\)

---

\(^{22}\) What if the neighbours do have such a habit, but my mother has no idea of it and just rashly assumes that they are there? Then she does not know. On the method infallibilist view, that result can be obtained in two ways. Either we have a close possibility in which she believes something false on the basis of the same method (e.g., that some other neighbours are there while they are not). Or we have a close possibility in which their habit are different. In the latter option, inductive knowledge will require that my mother’s method is somehow sensitive to the neighbours’ habits: for instance, that she would not make this inferences if they did not have the habit. The sensitivity
Let $m$ be the method that produces my mother’s belief that $(p)$ the neighbours are there. (See figure 5.5.)

At $w_1$, the method produces a true belief. As in the model for Chisholm’s sheep case (section 5.3.2), there is a world in which the method produces a false belief ($w_2$). But here that world is not a genuine possibility in $w_1$. The absence of neighbours while their light is on is not something that could have happened in the circumstances at hand. So $m$ is infallible at $w_1$, and and $p$ is known.

We can thus say that a method is inductive iff it is metaphysically fallible. If we want to call “strongly a priori” a piece of knowledge that is based on a non-inductive method, we get a vindication of the idea of contingent a priori knowledge: for instance, the method that would lead each subject to believe that she exists is metaphysically infallible.

Methods models thus allow us to draw an important distinction between two properties that the label “fallibility” fails to distinguish. A reliable method can be “fallible” in the sense that it is (a) fallible properly, that is, it produces false beliefs within the relevant set of worlds, (b) inductive, that is, it produces false beliefs outside of the relevant set of worlds.

An inductive method can thus be infallible in the proper sense. Fallibil-

Figure 5.5: Inductive knowledge

---

may amount to knowledge of their habits, but such knowledge can easily fall short of entailing that they are there tonight if the light is on tonight. Even in these more realistic settings, inductive knowledge boils down to local infallibility. (The account of inductive knowledge assumed here is externalist: whether a subject has inductive knowledge by a certain method depends on features of her environment she may not be aware of. See Armstrong, 1973, 157, 166–7, 206–8; Dretske, 1971, 2–4.)
ity in the first sense is incompatible with deductive closure and is likely to lead to Gettier cases; “fallibility” in the second sense (inductivity) is compatible with deductive closure and is more likely to deal with Gettier cases.

5.4 Methods models

All we have said so far assumed only a set of worlds, an alethic modality over them, and a set of propositions. That suffices to characterise methods, operations over them, their infallibility, and to give models of Gettier cases and inductive knowledge. To get a proper formal implementation of methods infallibilism, however, we now opt for a particular notion of proposition. What has been said so far nevertheless holds for other choices.

We take the simplest notion of proposition, namely a set of worlds. That creates trouble with so-called Frege cases, though we point out that even there our simplest method models have something interesting to say. We show how agents are specified as sets of methods, and how knowledge and belief are derived from those. We give a language and we state basic equivalence results with Scott-Montague neighbourhood models (Montague, 1968, 1970; Scott, 1970).

5.4.1 Propositions and the problem of Frege cases

Propositions as sets of worlds

**Definition 5.4.** Propositions are sets of worlds: $P = \mathcal{P}(W)$.

A proposition $p$ is true at a world $w$ iff $w \in p$.

The resulting methods models are an extension of neighbourhood models. See Appendix C.2.
Frege cases

The choice keeps our models simple and on familiar grounds, but comes at a cost. Call referential opacity cases cases in which we are tempted to say that an agent knows (or believes) that \( p \) but fails to know (believe) that \( q \), while \( p \) and \( q \) are true at exactly the same worlds.\(^{23}\) Well-known candidates are:

**Proper names** Alice knows that Hesperus shines, but she does not know that Phosphorus shines. ([Frege, 1980](#))

Pierre believes that Londres is pretty, but he does not believe that London is pretty. ([Kripke, 1979](#))

**Indexicals** David knows that that man’s pants [unwittingly pointing at himself in the mirror] are on fire, but he does not know that his pants are on fire. ([Kaplan, 1989](#), sec XVII, see also [Perry, 1979](#))

**Natural kinds terms** Saul knows that there is water in the glass, but does not know that there is \( \text{H}_2\text{O} \) in the glass. ([Kripke, 1980](#))

**Logical equivalents** Fred knows that \( p \) but does not know that \( p \lor (\neg p \rightarrow p) \).

Given (Definition [5.4](#)), if \( p \) and \( q \) are true at exactly the same world, \( p = q \), and thus \( p \) is an output of a method \( m \) iff \( q \) is. So it appears that simple methods models cannot represent referential opacity cases.

The matter is not so straightforward, however. Suppose that Alice is in a situation that we would be tempted to describe as follows: not knowing that Hesperus and Phosphorus are the same planet, she believes that Hesperus shines but not that Phosphorus shines. Let \( m^H \) be a constant for the method through which she believes that Hesperus shines (when she does), and \( m^p \) a constant for the method through which she believes that Phosphorus shines (when she does). (We introduce the full language in section [5.4.3](#).) In our language, we represent Alice’s situation as follows:

\[
Bm^H p \land \neg Bm^H \neg p \land Bm^P \neg p \land \neg Bm ^ P p
\]

---

23. The terminology comes from [Quine (1961)](#).
where \( p \) is a term for the proposition that Venus shines. In other terms: by method \( m^H \), she believes the proposition to be true and does not believe it to be false; by method \( m^P \), she believes it to be false and does not believe it to be true. That representation at least avoids a contradictory statement that \( Bp \land \neg Bp \). Moreover, it shows that the notion of method can at least partly capture Frege’s elusive notion of “mode of presentation”: a belief that \( p \) is a Hesperus-belief if based on \( m^H \), and a Phosphorus-belief if based \( m^P \). And it is compatible with a direct-referentialist view of content.

However, the suggested account has its limits. It is likely that referential opacity phenomena arise within a single method. For instance, if I see the end of a ship by one window, and its other end by another, I may rationally come to believe that they are two distinct ships (Perry, 1979, 483). I may thus believe that that ship [pointing at one end] is identical to that ship [pointing at the same end] while also believing that that ship [pointing at the other end] is not identical to that ship [pointing at the first end again]. We may want to count the two beliefs as being issued by the same method. So we get a case in which:

\[
(Bm : p) \land (Bm : \neg p)
\]

where \( p \) is the proposition that that ship is identical to itself, and \( m \) the method by which I formed both beliefs. If \( m \) is combined with some Deduction method, we may even get:

\[
Bn : (p \land \neg p)
\]

for \( n = m^D \circ m \) (see section ?? on Deduction methods). Here we cannot represent the contradiction as a contradiction of beliefs based on different sources. Here we cannot represent the contradiction as a contradiction of beliefs based on different sources. An option would be to make methods as fine-grained as necessary to make sure that Frege cases can all be

24. See Kripke (1979, section III).
cashed out in terms of distinct methods, but that strikes me as an *ad hoc* move, and would render the notion of method less natural.

Some further enrichment of the models thus appears needed to fully take into account referential opacity phenomena. One option is to distinguish propositions that are true at the same possible worlds: this can be done by introducing impossible worlds.\(^{25}\) Another option is to take the outputs of methods to be sentence-like structures. Philosophically, that option corresponds equally to (a) a language-of-thought view, (b) a Frege-style view in which propositions are structured and not reducible to extensions, (c) a view of belief as a ternary relation between subjects, modes of presentations (represented by sentences), and coarse propositions.\(^{26}\)

### 5.4.2 Frames and agents

Our frames are given by a set of worlds, an accessibility relation over them, and a set of methods for the agent. (In the multi-agent case, a set of methods can be introduced for each agent.) Propositions and the space of methods are themselves defined from the set of worlds:

**Definition 5.5.** A *methods frame* \(\mathfrak{F}\) is a triple \(\langle W, M^B, R \rangle\) where:

\(^{25}\) We introduce a set of impossible worlds \(I\) and redefine the set of propositions as \(P = \mathcal{P}(W \cup I)\). Premises and conclusions of methods are now taken from this extended set; however, they can remain functions from possible worlds. At impossible worlds, valuation is not compositional: any arbitrary set of formulas can hold — impossible worlds can be modelled as such sets. For two propositional constants \(p_1\) and \(p_2\) (standing for “Hesperus shines” and “Phosphorus shines”, for instance) that hold at exactly the same possible worlds in a model, we have an impossible world \(w^*\) such that \(w^* \in p_1\) and \(w^* \notin p_2\), so that \(p_1 \neq p_2\) and we can have \(p_1 \in m(w)\) and \(p_2 \notin m(w)\) for some method \(m\) and possible world \(w\).

\(^{26}\) We can take the premises and conclusions of methods to be *formulas* of our language, for instance — similar strategies appear in Awareness semantics (Fagin and Halpern, 1988) and Fitting’s semantics for the Logic of Proofs (Fitting, 2005). (Though it should be noted that the technique is philosophically unsatisfactory, since it makes things look as though the mental states of a subject were dependent on the language of the ascriber.) However, the resulting semantics is prone to self-referential paradoxes, and methods may have to be typed in order to avoid them.
\( M^B \subseteq M \) is the set of basic methods of the agent, where \( M = W \rightarrow (\mathcal{P}(P) \rightarrow \mathcal{P}(P)) \) with propositions as sets of worlds: \( P = \mathcal{P}(W). \)

\( R \subseteq W \times W \) is a reflexive accessibility relation over worlds.

A transitive methods frame is a methods frame in which \( R \) is transitive.

The primitives of our models are just worlds, a set of basic methods for each agent, and a background alethic modality given by \( \langle W, R \rangle \) that will be used to characterise infallibility. The rest is derived as follows.

**Definition 5.6.** The agent’s method set is \( M = M^B \circ +. \)

An agent is represented by sets of methods. We assume an agent free of modular or computational limitations: the set of her methods is closed under composition and union. As we noted in section 5.2.1, bounded agents can be modelled by putting restrictions on building \( M \) out of \( M^B \), and more sophisticated representations of agents could have their method set changing through time; the latter rather may be represented as functions from times to sets of methods.

We define a set of infallible methods at each world:

**Definition 5.7.** \( M^I(w) \) is the set of infallible methods at \( w \): \( m \in M^I(w) \) iff \( \forall w', wRw': \forall p (p \in m(w') \rightarrow w' \in p) \).

**Lemma 5.1.** If \( \mathcal{R} \) is a transitive methods frame, then for any \( m, w, w' \) such that \( wRw' \), if \( m \in M^I(w) \) then for any \( w' \) such that \( wRw' \), \( m \in M^I(w') \).

**Proof.** Let \( m, w \) be such that \( m \in M^I(w) \), and \( w' \) such that \( wRw' \). Let \( w'' \) be any world such that \( w'Rw'' \). By the transitivity of \( R \), \( wRw'' \). Since \( m \in M^I(w) \), by Definition 5.7, \( \forall p (p \in m(w'') \rightarrow w'' \in p) \). By Definition 5.7 again, \( m \in M^I(w') \).

Lemma 5.1 is crucial to the derivation of knowledge of one’s knowledge (section 5.16).

---

27. Cf. sections 5.2.1 and 5.4.1.
28. Cf. section 5.2.2.
Finally we define a range of functions $W \times M \rightarrow \mathcal{P}(P)$ and $W \rightarrow \mathcal{P}(P)$ for belief and knowledge. "B" and "K" each refer to two functions, but the ambiguity is convenient and their arguments always disambiguate:

**Definition 5.8.** $B : (m, w) \mapsto \{p : m \in M \land p \in m(w)\}$ gives the agent’s beliefs on the basis of $m$.

$K : (m, w) \mapsto \{p : p \in B(m, w) \land m \in M^I(w)\}$ gives the agent’s knowledge on the basis of $m$.

$B : w \mapsto \{p : \exists m(p \in B(m, w))\}$ gives the agent’s beliefs simpliciter.

$K : w \mapsto \{p : \exists m(p \in K(m, w))\}$ gives the agent’s knowledge simpliciter.

Beliefs are just the outputs of the agent’s methods on the basis of no premises. Knowledge is belief on an infallible basis. $B(m, w)$ is the set of propositions believed on the basis of $m$ at $w$, and $B(w)$ is the set of propositions believed at $w$, and similarly for knowledge. $B(w)$ and $K(w)$, as well as the functions $w \mapsto B(m, w)$ and $w \mapsto K(m, w)$ for a given $m$, are neighbourhood functions (see Appendix C.2).

The definitions of belief and knowledge simpliciter are not entirely innocuous. For all we have said, an agent may be such that one of her methods unconditionally outputs $p$ and another $\neg p$. The definition implies that such an agent both believes $p$ and not-$p$. Some may want to resist that; one may want to say for instance that an agent believes $p$ iff one of her methods outputs $p$ and no other outputs $\neg p$. However, our existentially quantified definition is by far the simplest; alternative options create holistic constraints on belief and knowledge that would prevent from getting fully general theorems such as $Kp \rightarrow KKp$. I do not see the fact that we rely on this choice as an important liability, though; for one lesson of methods models is that deep generalisations about knowledge should be stated in terms of *based* belief and knowledge rather than in terms of belief and knowledge *simpliciter*. 
5.4.3 Language

Syntax

Definition 5.9. Let \( M = \{m, n, \ldots\} \) be a set of methods constants. The set of methods terms is given by the grammar:

\[
\mu ::= m|\nu|\mu \circ \nu
\]

\( \mathcal{L} \) is the set of formulas given by:

\[
\phi ::= p|T|\neg\phi|\psi|\phi \land \psi|\phi \rightarrow \psi|\phi \leftrightarrow \psi|B\mu : \psi|K\mu : \phi|B\psi|K\phi|\Box \phi
\]

where \( P = \{p, q, r, \ldots\} \) is a set of propositional constants.

We introduce two \( B \) and \( K \) operators, one for methods-relative belief and knowledge and the other for belief and knowledge simpliciter. (Introducing them separately removes the need for quantification over methods.) \( \Box \) will express the background alethic modality.

By convention, \( B\mu : \phi \rightarrow \psi \) as \( (B\mu : \phi) \rightarrow \psi \) and not as \( B\mu : (\phi \rightarrow \psi) \).

Semantics

Definition 5.10. Let \( \mathfrak{M} =< \mathfrak{I}, V > \) be a model where \( V : P \cup M \rightarrow P \cup M \) is a valuation function that assigns a proposition to each propositional constant and a method to each method constant. We define \( \llbracket \cdot \rrbracket_{\mathfrak{M}} \):

\[
\llbracket p \rrbracket_{\mathfrak{M}} = V(p), \\
\llbracket m \rrbracket_{\mathfrak{M}} = V(m), \\
\llbracket \mu + \nu \rrbracket_{\mathfrak{M}} = \llbracket \mu \rrbracket_{\mathfrak{M}} + \llbracket \nu \rrbracket_{\mathfrak{M}}, \\
\llbracket \mu \circ \nu \rrbracket_{\mathfrak{M}} = \llbracket \mu \rrbracket_{\mathfrak{M}} \circ \llbracket \nu \rrbracket_{\mathfrak{M}}, \\
\llbracket T \rrbracket_{\mathfrak{M}} = W, \\
\llbracket \neg \phi \rrbracket_{\mathfrak{M}} = W \setminus \llbracket \phi \rrbracket_{\mathfrak{M}} \\
\llbracket \phi \lor \psi \rrbracket_{\mathfrak{M}} = \llbracket \phi \rrbracket_{\mathfrak{M}} \cup \llbracket \psi \rrbracket_{\mathfrak{M}}
\]

(and as usual for other logical connectives)

\[
\llbracket B\mu : \phi \rrbracket_{\mathfrak{M}} = \{w : \llbracket \phi \rrbracket_{\mathfrak{M}} \in B(\llbracket \mu \rrbracket_{\mathfrak{M}}, w)\},
\]
\[
\begin{align*}
\lbrack \text{K}_\mu : \phi \rbrack^{\text{3R}} &= \{ w : \lbrack \phi \rbrack^{\text{3R}} \in K(\lbrack \mu \rbrack^{\text{3R}}, w) \} \\
\lbrack \text{B}_\phi \rbrack^{\text{3R}} &= \{ w : \lbrack \phi \rbrack^{\text{3R}} \in B(w) \} \\
\lbrack \text{K}_\phi \rbrack^{\text{3R}} &= \{ w : \lbrack \phi \rbrack^{\text{3R}} \in K(w) \} \\
\lbrack \Box \phi \rbrack^{\text{3R}} &= \{ w : \forall w'(wRw' \rightarrow w' \in \lbrack \phi \rbrack^{\text{3R}}) \}.
\end{align*}
\]

Truth. \( \models_{w}^{\text{3R}} \phi \) iff \( w \in \lbrack \phi \rbrack^{\text{3R}} \).

Validity. \( \models_{w}^{\text{3R}} \phi \) iff for any \( w, \models_{w}^{\text{3R}} \phi \).

The clause for \( \Box \phi \) is familiar from Kripke models. The clauses for \( \text{B}_\phi \) and \( \text{K}_\phi \) are familiar from neighbourhood models, and it is easy to see that the clauses for \( \text{B}_\mu : \phi \) and \( \text{K}_\mu : \phi \) pick up the neighbourhood function corresponding to the unconditional outputs method \( m \) designated by \( \mu \), namely: \( w \mapsto m(w) \), provided that it is one of the agent’s method (belief case) and that it is infallible (knowledge case).

In Appendix C.2 we compare neighbourhood models and methods models in more detail. We show that for any neighbourhood model for \( \text{B} \), there is an equivalent method model for \( \text{B} \), and conversely, and that for any neighbourhood model for \( \text{K} \) in which \( \text{K}_\phi \rightarrow \phi \) is valid, there is an equivalent method model for \( \text{K} \), and conversely. But we also argue that methods models are more explanatory than neighbourhood ones, because they allow us to derive the modal axioms from the structure of an agent’s methods.

We check that belief and knowledge on a basis entail belief and knowledge *simpliciter*:

**Corollary 5.1.** For any \( \mathfrak{M}, w, \models_{w}^{\text{3R}} \text{B}_\mu : \phi \rightarrow \text{B}_\phi \) and \( \models_{w}^{\text{3R}} \text{K}_\mu : \phi \rightarrow \text{K}_\phi \) for any \( \mu, \phi \).

**Proof.** From Definition 5.10, \( \models_{w}^{\text{3R}} \text{B}_\mu : \phi \) iff there is an \( m \in M \) such that \( \lbrack \phi \rbrack^{\text{3R}} \in m(w) \). By Definition 5.8, \( \lbrack \phi \rbrack^{\text{3R}} \in B(w) \), and by Definition 5.10 again, \( \models_{w}^{\text{3R}} \text{B}_\phi \). And analogously for \( \text{K} \). \( \square \)

The methods algebra of Appendix C.1 and the semantics give us a series of equivalences:

---

29. Alternatively, it is straightforward to introduce quantification over methods in the language. We can then define \( \text{B}_\phi \) as \( \exists \mu(\text{B}_\mu : \phi) \) and similarly for \( \text{K}_\phi \).
Corollary 5.2. The following are valid in any method model $\mathcal{M}$:

- $\mathbf{B}(\mu + \nu + \rho : \phi) \leftrightarrow \mathbf{B}\mu + (\nu + \rho) : \phi$,
- $\mathbf{B}\mu + \nu : \phi \leftrightarrow \mathbf{B}\nu + \mu : \phi$,
- $\mathbf{B}\mu + \mu : \phi \leftrightarrow \mathbf{B}\mu : \phi$,
- $\mathbf{B}(\mu \circ \nu) \circ \rho : \phi \leftrightarrow \mathbf{B}\mu \circ (\nu \circ \rho) : \phi$,
- $\mathbf{B}(\mu + \nu) \circ \rho \leftrightarrow \mathbf{B}(\mu \circ \rho) + (\nu \circ \rho) : \phi$,

and similarly for $\mathbf{K}$, for any $\mu, \nu, \rho, \phi$.

For some methods model $\mathcal{M}$, $\not\models^{\mathcal{M}} \mathbf{B}\mu \circ (\nu + \rho) : \phi \leftrightarrow \mathbf{B}(\mu \circ \nu) + (\mu \circ \rho) : \phi$,

and similarly for $\mathbf{K}$, for some $\mu, \nu, \rho, \phi$.

5.5 Results

The consequences of methods models fall within four groups, corresponding to different idealisations of agents:

1. For any agent: knowledge entails belief and truth (subjectivity and factivity of knowledge). That is a welcome result, since these are the only two (quasi-)uncontroversial facts about knowledge. Moreover, we have referential transparency: if $p$ and $q$ are true exactly at the same worlds, $p$ is known iff $q$ is. That is a limitation of the simpler models, as we noted (section 5.4.1).

2. For perfect reasoners, who have specific methods to believe all logical truths and all the logical consequences of what they believe: deductive closure. With unbounded resources, this validates the logical omniscience axiom $\mathbf{K} \phi \rightarrow \mathbf{K}\phi$.

3. For perfect introspecters and perfectly confident introspecters, who have methods to ensure that they believe that they believe $p$ whenever they believe $p$ (positive psychological introspection), that they believe that they do not believe $p$ when they do not (negative psychological introspection), that they believe that they know $p$ whenever they believe $p$ (positive confident introspection) and that they believe that they do not know $p$ when they do not believe it (negative confident introspection): self-knowledge ($\mathbf{B}\phi \leftrightarrow \mathbf{K}\phi$), partial negative
epistemic introspection \(\neg B\phi \rightarrow K\neg K\phi\) and, if possible possibilities are possible, epistemic positive introspection or axiom \(4\) \(K\phi \rightarrow KK\phi\).

4. For excellent agents, whose methods are all infallible: believing is knowing \((B\phi \leftrightarrow K\phi)\) and epistemic negative introspection or axiom \(5\) \((\neg K\phi \rightarrow K\neg K\phi)\).

With pure reasoners, we get a normal modal logic \(K\) for the belief simiplicity operator and \(KT\) for the knowledge simplicity operator, and putting all idealisations together, we get a standard \(S5\) system for both. This gives us an equivalence between those models and standard Hintikka models for the subpart of \(L\) that is free of method terms, and shows that methods models can be as powerful as the standard ones.

But as the summary indicates, the idealisations which we use to derive our results are natural idealisations of the psychology of an agent. (For positive introspection, we need also an assumption on the structure of possibilities.) They are more intuitive than direct judgement on the \(S5\) axioms or on formal constraints on accessibility relations (transitivity, euclideanity). Correspondingly, they give us a better understanding of why and when the various axioms of standard epistemic logic hold.\(^{30}\)

### 5.5.1 Subjectivity, factivity, and referential transparency

Referential transparency: the rule of equivalence

**Theorem 5.1.** Referential transparency \((E_{BK})\). If \(\models_{S} \phi \leftrightarrow \psi\), then \(\models_{S} B\phi \leftrightarrow B\psi\) and \(\models_{S} K\psi \leftrightarrow K\phi\), for any methods model \(M\).

**Proof.** Suppose \(\models_{S} \phi \leftrightarrow \psi\). We have \([\phi]^{S} = [\psi]^{S}\), so \([B\phi]^{S} = \{w : [\phi]^{S} \in B(w)\} = \{w : [\psi]^{S} \in B(w)\} = [B\psi]^{S}\), and similarly for \(K\). \(\Box\)

30. The properties of the epistemic accessibility relation in standard epistemic logic can be naturally interpreted as indistinguishability relations. But this hides an important ambiguity. Indistinguishability can be understood as inability to know the difference: \(w\) is indistinguishable from \(w'\) to one iff in \(w\) one cannot know that \(w\) is different from \(w'\). Or indistinguishability can be understood as sameness of internal state: \(w\) is indistinguishable from \(w'\) to one iff in \(w\) one is in the same internal state as in \(w'\). The first reading is Hintikka’s (2007) and the second is roughly Lewis’ (1996). Each has problematic consequences, as we noted in section 5.1. (Thanks to an anonymous referee here.)
Referential transparency is the rule of equivalence of classical modal logic. It is known to determine the class of all neighbourhood frames. \(^{31}\)

Thus by Theorem C.5 the schema determines methods frames for the \(\mathcal{B}\) operator.

Referential Transparency is a consequence of our choice of modelling propositions as sets of possible worlds. It ensures that our models are classical and similar to neighbourhood models. However, in doxastic and epistemic terms, that means that belief and knowledge are referentially transparent in the sense discussed above. This is a problematic consequence, as we said (section 5.4.1).

**Factivity and subjectivity**

**Theorem 5.2.** Subjectivity (S). \(\models^{30} K\phi \rightarrow B\phi\) for any methods model \(\mathcal{M}\).

*Proof.* Evident from Definitions 5.8 and 5.10. \(\Box\)

The theorem states that knowledge is subjective, in the sense that knowledge is in part a matter of the agent’s psychology, namely, his beliefs. \(^{32}\)

**Theorem 5.3.** Factivity (T\(_K\)). \(\models^{30} K\phi \rightarrow \phi\) for any methods model \(\mathcal{M}\).

*Proof.* Evident from the fact that \(R\) is reflexive and Definition 5.7, Definition 5.8 and Definition 5.10. \(\Box\)

Factivity entails that knowledge is consistent, or axiom D: \(K\phi \rightarrow \neg K\neg\phi\). (Assume \(K\phi\); by factivity, \(\phi\), so \(\neg\neg\phi\), and by factivity again, \(\neg K\neg\phi\).)

---


32. It should thus be clear that when I say that knowledge is subjective I do not mean at all that it is “subjective” in the sense in which matters of taste are said to be “subjective”, i.e. merely a matter of people’s opinions or preferences. I just mean that it requires a state of mind. That is in contrast with the notion of “implicit knowledge” that Hintikka models are often taken to formalise, which does not require an agent be in any sense aware of things it “knows”. 
It is known from neighbourhood semantics that $E_{BK}$ and $T_K$ determine truthful neighbourhood models. By Theorem C.6 the schemas determine methods models with respect to the $K$ operator.

Note that we need two things to derive Factivity. First we need the reflexivity of $R$, that is, $\Box$ should be a modality that itself satisfies $T$:

**Theorem 5.4.** Alethic necessity ($T_\Box$). $\models^{\forall w} \Box \phi \rightarrow \phi$ for any method model $\mathcal{M}$.

*Proof.* From the reflexivity of $R$ and Definition 5.10. \qed

This reflects the fact that $\Box$ is intended as an alethic modality. Second, we need strict infallibility, i.e. that the methods in $M^I(w)$ are such that all their outputs are true.

Suppose we try to put a weaker condition on knowledge; for instance, we include in $M^I(w)$ all the methods that are highly reliable at $w$ (see section 5.3.3). Factivity can no longer be derived: $p \in m(w)$ and $m \in M^I(w)$ do not entail that $p$ is true at $w$, since some of the unconditional outputs of $m$ may be false at $w$. Truth must then be added as a separate condition on knowledge: $p \in K(w)$ iff there is a $m$ such that $p \in m(w)$, $m \in M^I(w)$ and $w \in p$. This is in essence the “justified-true-belief” analysis of knowledge, and it is open to Gettier counterexamples because its two conjuncts ($m$ is a reliable/adequate method, and $p$ is true) can be simultaneously satisfied by coincidence.

**Failure of logical omniscience and introspection**

None of the other axioms of modal logic are valid.

**Theorem 5.5.** Each of $M$, $C$, $K$, $N$, $4$, $5$ for $\mathcal{B}$ and $K$ fails in some methods model $\mathcal{M}$, and $D$ and $T$ for $\mathcal{B}$ fail in some methods model $\mathcal{M}$.

---

33. A neighbourhood model $(W, N)$ is truthful iff $w \in \bigcap N(w)$ for any $w$. See Appendix C.2.
5.5. Results

\[ D_B. \quad B\phi \rightarrow \neg B\neg \phi \]
\[ T_B. \quad B\phi \rightarrow \phi \]
\[ M_B. \quad B(\phi \land \psi) \rightarrow (B\phi \land B\psi) \quad M_K. \quad K(\phi \land \psi) \rightarrow (K\phi \land K\psi) \]
\[ C_B. \quad (B\phi \land B\psi) \rightarrow B(\phi \land \psi) \quad C_K. \quad (K\phi \land K\psi) \rightarrow K(\phi \land \psi) \]
\[ K_B. \quad B(\phi \rightarrow \psi) \rightarrow (B\phi \rightarrow B\psi) \quad K_K. \quad K(\phi \rightarrow \psi) \rightarrow (K\phi \rightarrow K\psi) \]
\[ N_B. \quad B\top \quad N_K. \quad K\top \]
\[ 4_B. \quad B\phi \rightarrow BB\phi \quad 4_K. \quad K\phi \rightarrow KK\phi \]
\[ 5_B. \quad \neg B\phi \rightarrow B\neg B\phi \quad 5_K. \quad \neg K\phi \rightarrow K\neg K\phi \]

**Proof.** Neighbourhood models that invalidate each schemas are known. By Theorem C.5 they can be used to show that none of the B schema are valid in methods models. Moreover, it is easy to construct such countermodels as truthful neighbourhood models, so that by theorem C.6 we have counterexamples to the K schemas. \[ \square \]

In appendix C.3, we give illustrations of violations of M, N, K, and 4. In particular, the violation of K illustrates our Watson case (see p. 421 and p. 227 for the case.).

5.5.2 Perfect reasoning

The first interesting class of methods models is that of perfect reasoners. Intuitively, an agent is a perfect reasoner if she believes all logical truths, and deduces all logical consequences of what she believes. To model such agents, we define the two following methods:

**Definition 5.11.** The Pure Reason method \( m^R \) is the method such that
\[ m^R(w, \pi) = \{W\} \]
for any \( w, \pi \).\[34\]

For any model \( \mathfrak{M} \), we write \( m^R \) the constant such that \( \mathfrak{M}[m^R] = m^R \).

That is, Pure Reason outputs the tautology given any set of premises, including the empty set.

\[34 \] Thus defined, Pure Reason entails that the agent exists at any world. To avoid this, one could instead use a Conditional Pure Reason method: \( m^{CR}(w, \pi) = \{W\} \) for any \( w, \pi \neq \emptyset \). Conditional Pure Reason outputs the tautology only if the agent has another belief. The resulting schema are \( \models^{5R} B\phi \rightarrow B\top \) and \( \models^{5R} B\phi \rightarrow K\top \) for any \( \phi \).
Definition 5.12. The Multi-Premise Deduction method \( m^D \) is the method such that 
\[
\{ p : \exists q, r \in \pi(q \cap r \subseteq p) \} \text{ for any } w, \pi.
\]

For any model \( \mathfrak{M} \), we write \( m^D \) the constant such that 
\[ [m^D]^{\mathfrak{M}} = m^D. \]

That is, Deduction maps a set of premises to all logical consequences of pairs of premises. (Note that it outputs nothing on the basis of the empty set of premises.)\(^{35}\)

We call a perfect reasoner model a methods model such that 
\[ m^R, m^D \in M. \]

Theorem 5.6. Knowledge of logic (\( N_B \) and \( N_K \)). For any perfect reasoner model \( \mathfrak{M} \), 
\[ \models^{\mathfrak{M}} B^R : \top \text{ and } \models^{\mathfrak{M}} K^R : \top. \]

Proof. We first prove that at any world, the agent believes the tautology on the basis of Pure Reason; we then prove that Pure Reason is infallible at any world.

Let \( w \) be any world in a perfect reasoner model \( \mathfrak{M} \). By Definition 5.11, \( W \in m^R(w) \). By Definition 5.8, \( W \in B(w) \). By Definition 5.10, \( \models^{\mathfrak{M}} B^T \).

Furthermore, by Definition 5.11, for any \( p, w \), if \( p \in m^R(w) \) then \( p = W \), so \( w \in p \). Thus by Definition 5.7, \( m^R \in M^I(w) \). By Definition 5.8, \( W \in K(w) \), and by Definition 5.10, \( \models^{\mathfrak{M}} K^T \). \( \Box \)

Theorem 5.7. Deductive closure. For any perfect reasoner model \( \mathfrak{M} \):
\[ \models^{\mathfrak{M}} B^R : (\phi \rightarrow \psi) \rightarrow (B^R : \phi \rightarrow B^D \circ (\mu + \nu) : \psi) \]

\(^{35}\) In particular, given any premisse, Deduction will output the tautology. But that does not make Pure Reason redundant. If all pure non-inferential methods of the agent are fallible, then composing Deduction with them cannot yield knowledge, since the resulting composed method is fallible as well. Adding Pure Reason to the method set of such an agent enables knowledge of tautologies. (Thanks to an anonymous referee here.)

\(^{36}\) Note that an agent may be a perfect reasoner without having Pure Reason and Deduction in its basic set \( M^R \). To illustrate, let \( m, n \in M^R \) be such that, for some \( p \subseteq W \), \( m(w, \pi) = \{ W \} \) if \( w \in p \) and \( \emptyset \) otherwise, and \( n(w, \pi) = \{ W \} \) if \( w \notin p \) and \( \emptyset \) otherwise. (\( m \) outputs the tautology at \( p \) worlds, \( n \) outputs the tautology at not-\( p \) worlds.) We have \( m + n = m^R \). So if \( M \) is the union and composition closure of \( M^R \), \( m + n \in M \), and the corresponding model is a perfect reasoner model provided that \( m^D \in M \) as well, even if \( m^K \notin M^R \).
\[ \models_{\mathcal{W}} K\mu : (\phi \rightarrow \psi) \rightarrow (K\nu : \phi \rightarrow Km^D \circ (\mu + \nu) : \psi) \]

for any \( \mu, \nu, \phi, \psi \). Thus by Corollary 5.1, \( \models_{\mathcal{W}} B(\phi \rightarrow \psi) \rightarrow (B\phi \rightarrow B\psi) \) and \( \models_{\mathcal{W}} K(\phi \rightarrow \psi) \rightarrow (K\phi \rightarrow K\psi) \) for any \( \phi, \psi \).

Proof. We show that whenever \( p \rightarrow q \) and \( p \) are believed on the basis of \( m \) and \( n \) respectively, \( q \) is believed by \( m^D \circ (m + n) \). We additionally show that \( m^D \circ (m + n) \) is infallible at any world.

Let \( \mathcal{M}, w \) be a perfect reasoner model and a world such that \( \models_{\mathcal{W}} B\mu : (\phi \rightarrow \psi) \land B\nu : \phi \) for some \( \mu, \nu, \phi, \psi \). By Definition 5.10 and Definition 5.8, there are \( m, n \) and \( p, q \) such that \( \llbracket \mu \rrbracket_{\mathcal{W}} = m \), \( \llbracket \nu \rrbracket_{\mathcal{W}} = n \), \( \llbracket \phi \rrbracket_{\mathcal{W}} = p \), \( \llbracket \psi \rrbracket_{\mathcal{W}} = q \), \( (Wp) \cup q \in m(w) \) and \( p \in n(w) \) and \( m, n \in M \). By Definition 5.2, \( (Wp) \cup q, p \in (m + n)(w) \). Since \( ((Wp) \cup q) \cap p \subseteq q \), by Definition 5.12 we have \( q \in m^D \circ (m + n)(w) \). By Definition 5.10, \( \models_{\mathcal{W}} Bm^D \circ (\mu + \nu) : \psi \), which completes the proof of \( K_{B} \).

Now suppose \( \models_{\mathcal{W}} K\mu : (\phi \rightarrow \psi) \land K\nu : \phi \) for a perfect reasoner model \( \mathcal{M} \), a world \( w \), and some \( \phi, \psi \). The situation is as before with, additionally, \( m, n \in M'(w) \) (Definitions 5.10 and 5.8). We show that \( m^D \) preserves infallibility: \( m^D \circ (m + n) \in M'(w) \). Suppose, for some \( w', p' : wRw' \) and \( p' \in (m^D \circ (m + n))(w) \). By Definition 5.12, there are \( q', r' \in (m + n)(w) \) such that \( p' \subseteq q' \cap r' \). Since by assumption \( m, n \in M'(w) \) and \( wRw' \), by Definition 5.7 we have \( w' \in q' \) and \( w' \in r' \), and since \( q' \cap r' \subseteq p' \), it follows that \( w' \in p' \). So by Definition 5.7 again, \( m^D \circ (m + n) \in M'(w) \). We show as before that \( m^D \circ (m + n) \in M \) and \( q \in (m^D \circ (m + n))(w) \), so by Definition 5.10 \( \models_{\mathcal{W}} Km^D \circ (\mu + \nu) : \psi \), which completes the proof of \( K_{K} \). \( \square \)

The proof of \( K_{K} \) relies on two things. First, for any methods \( m, n \), the agent has a union method \( m + n \) that outputs the union of the original outputs of \( m \) and \( n \). This means that the agent “puts together” the result of any two methods. (Note that this does not mean that she believes the conjunction of original outputs.) Second, given any method \( m \), \( m^D \circ m \) outputs all the logical consequents of any two conclusions reached by \( m \). We show that \( m^D \circ m \) is infallible if \( m \) is, and that if \( p, q \) are unconditional outputs of \( m \), then \( m^D \circ m \) outputs all logical consequents of \( p \land q \).
It is easy to see that $m^D \circ m$ will output the consequents of any two premises given my $m$, $m^D \circ m^D \circ m$ the consequents of any three premises, $m^D \circ m^D \circ m^D \circ m$ the consequents of any four premises, and so on. Correspondingly, we can limit the number of consequents the agent is able to reach by putting limits on the number of repeated applications of Deduction she can make, and we can model a dynamic process of reasoning by indexing those limits to time.

Methods models thus allow us to draw a distinction between deductive closure proper and logical omniscience. If an agent is limited on the number of $m^D$ steps that she can reach, she will not be logically omniscient. But still, any consequence of what she knows that she does deduce will be knowledge, since $m^D$ preserves infallibility. The idea that any agent knows all the consequences she does deduce can be somewhat indirectly captured by the following theorem:

**Theorem 5.8.** Deduction preserves knowledge. For any methods model $\mathcal{M}$:

$$\models^\mathcal{M} K_\mu : \phi \rightarrow (Bm^D \circ \mu : \psi \rightarrow Km^D \circ \mu : \psi)$$

(If anything is known on the basis of $\mu$, then anything believed on the basis of deduction from $\mu$ is known).

**Proof.** Let $\mathcal{M}, w$ be such that $\models^\mathcal{M} w K_\mu : \phi$ for some $\phi$. By Definitions 5.10 and 5.8, $[\mu]^R_\mathcal{M} \in M'(w)$. We prove as in Theorem 5.7 that deduction preserves infallibility: if $[\mu]^R_\mathcal{M} \in M'(w)$, $m^D \circ [\mu]^R_\mathcal{M} \in M'(w)$. Now suppose that $\models^\mathcal{M} w Bm^D \circ \mu : \psi$ for some $\psi$. Since $[m^D \circ \mu]^R_\mathcal{M} = m^D \circ [\mu]^R_\mathcal{M} \in M'(w)$, $\models^\mathcal{M} w Km^D \circ \mu : \psi$. $\square$

The theorem has no equivalent in a language without methods terms. (The closest we can formulate is $K_\phi \rightarrow ((\Box (\phi \rightarrow \psi) \land B \psi) \rightarrow K \psi)$, which is counterexamplified if the agent believes $\psi$ from some other reasons than $\phi$, as in our Watson case. Note that $K_\phi \rightarrow (((\phi \rightarrow \psi) \land B \psi) \rightarrow K \psi)$ reduces to $K_\phi \rightarrow ((\psi \land B \psi) \rightarrow K \psi)$, which, barring bizarre cases, holds only for excellent agents: see Definition 5.18, p. 267.)
Further exploration of the deductive aspects of methods models are made in Appendix C.4:

1. Axioms \( M \) and \( C \) for belief and knowledge follow from axiom \( K \). But it is also possible to get them separately, by defining a Single-Premise Deduction method \( m^{SD} \) (for \( M \)) and a Conjunctive Deduction method \( m^{CD} \) (for \( C \)). The relation between \( m^D \), \( m^{SD} \) and \( m^{CD} \) is: 
\[
m^D = m^{SD} \circ m^{CD} \text{ (Corollary C.1).}
\]
2. These results are correlated to topological properties of sets of sets, as in neighbourhood semantics.
3. The \( B \) and \( K \) versions of the \( K \) axiom are mutually independent, and similarly for \( M, C, N \).
4. Having Pure Reason and Deduction is sufficient to satisfy the \( N \) and \( K \) axioms, respectively, but not necessary. So we have not characterised the full class of methods models that validates the axioms. However, we argue that that is not a defect: some agents might well satisfy the axioms without having specific methods for doing so, but that is then a coincidence, so to speak. The important regularities about knowledge are the methods-relative ones, not the ones stated in terms of belief and knowledge simpliciter.

### 5.5.3 Consistency

Pure Reason and Deduction do not guarantee than an agent is consistent: that is, the \( D_B \) axiom for belief (\( B\phi \rightarrow \neg B\neg \phi \)) may fail. (The \( D_K \) axiom for knowledge is of course guaranteed by the factivity of knowledge, Theorem 5.3.) Can we define a method to validate axiom \( D_B \)? No. And that is an intuitive result: avoiding contradictions among one’s beliefs is not a matter of forming one’s beliefs, but rather of revising them. As long as we have not defined methods for belief revision — for instance, functions from a method set to another —, we cannot define a method that ensures consistency. We can at most give a (trivial) constraint to satisfy consistency:
Definition 5.13. A consistent agent model is a methods model such that for $w, \bigcap B(w) \neq \emptyset$.

Theorem 5.9. If $\mathcal{M}$ is a consistent agent model, $(D_B) \vDash B\phi \rightarrow \neg B\neg\phi$ for any $\phi$.

Proof. Suppose $\mathcal{M}$ is a consistent agent model and $w$ a world such that $\vDash_B \phi$ for some $\phi$. By Definition 5.10 there is a $p$ such that $\llbracket \phi \rrbracket^{\mathcal{M}} = p$ and $p \in B(w)$. By Definition 5.13 it follows that $(W \setminus p) \not\vDash_B \neg \phi$. □

5.5.4 Perfect introspection and perfect confidence

The next two interesting classes of models are that of perfect introspecters and confident introspecters. Intuitively, an agent is a perfect introspecter if whenever she believes something, she believes that she does, and whenever she does not believe something, she believes that she does not. An agent is a confident introspecter if whenever she believes something, she believes that she knows it, and whenever she does not believe something, she believes that she does not know it.

Some abbreviations will be useful:

Definition 5.14. $b_p := \{w : p \in B(w)\}$,
$-b_p := \{w : p \not\in B(w)\} = W\setminus b_p$,
$k_p := \{w : p \in K(w)\}$,
$-k_p := \{w : p \not\in K(w)\} = W\setminus k_p$.

Perfect introspection: self-knowledge

Definition 5.15. A perfect introspecter model is a methods model for which there are positive and negative Introspection methods $m^p, m^n \in M$, such that $m^p(w, \pi) = \{b_p : p \in \pi\}$ for any $w, \pi$, and $m^n(w, \pi) = \{-b_p : p \not\in B(w)\}$ for any $w, \pi$.\footnote{Positive introspection does not have unconditional outputs: if $\pi = \emptyset$, $m^p(w, \pi) = \emptyset$ for any $w$. By contrast, negative introspection outputs propositions unconditionally.}

37. Positive introspection does not have unconditional outputs: if $\pi = \emptyset$, $m^p(w, \pi) = \emptyset$ for any $w$. By contrast, negative introspection outputs propositions unconditionally.
5.5. Results

Positive introspection methods and negative introspection methods are defined with respect to a given agent’s method set $M$. The definitions cannot uniquely characterise positive and negative introspection methods, because they are potentially self-referential: $b_p$ and $-b_p$ are defined in terms of $B(w)$, and the latter is defined in terms of the methods in $M$, which include $m^p$ and $m^\neg p$ themselves.

**Theorem 5.10.** Self-knowledge ($SK$). If $\mathcal{W}$ is a perfect introspector model, $\models^\mathcal{W}_w B\phi \rightarrow KB\phi$ and $\models^\mathcal{W}_w -B\phi \rightarrow K-B\phi$.

*Proof.* For positive self-knowledge, we show that whenever $p \in m(w)$ for some $m$, $b_p \in (m^p \circ m)(w)$ for some positive introspection method, and that $m^p \circ m$ is infallible.

Suppose $\mathcal{W}$ is a perfect introspector model and $w$ any world such that $\models^\mathcal{W}_w B\phi$. By Definitions 5.10 and 5.8, there are $p, m$ such that $\llbracket \phi \rrbracket^\mathcal{W}_w = p$, $m \in M$ and $p \in m(w)$. By Definition 5.15, there is an $m^p \in M$ such that $b_p \in (m^p \circ m)(w)$. Moreover, suppose that $p' \in (m^p \circ m)(w')$ for some $p', w'$. By Definitions 5.15 and 5.14, there is a $q'$ such that $p' = b_{q'}$ and $q' \in m(w')$ so that $w' \in b_{q'}$. So by Definition 5.7, $m^p \circ m \in M'(w')$. Finally, by Definitions 5.14 and 5.10, $b_p = [B\phi]^\mathcal{W}_w$. Since we have $m^p \circ m \in M$ (by $m \in M$ and Definition 5.6), $m^p \circ m \in M'(w)$, and $b_p \in (m^p \circ m)(w)$. By Definitions 5.8 and 5.10, $\models^\mathcal{W}_w KB\phi$.

For negative self-knowledge, we show that whenever $p \notin B(w)$, $-b_p \in m^\neg p(w)$ for some negative introspection method $m^\neg p$, and that $m^\neg p$ is infallible.

Suppose $\mathcal{W}$ is a perfect introspector model and $w$ any world such that $\models^\mathcal{W}_w -B\phi$. By Definitions 5.10 and 5.8, there is a $p$ such that $\llbracket \phi \rrbracket^\mathcal{W}_w = p$ and $p \notin B(w)$. By Definition 5.15, there is a $m^\neg p \in M$ such that $-b_p \in m^\neg p(w)$. Moreover, suppose that $p' \in m^\neg p(w')$ for some $p', w'$. By Definition 5.15 and

Since it does so at every world in which the agent does not believe some proposition, it implies that the agent exists at every world. (Negative introspection would not output anything at a world in which the agent already believes every proposition, but the agent would obviously exist at that world as well.) To lift that assumption, we could use conditional negative introspection methods $m^c\neg p$ such that $m^c\neg p(w, \pi) = \{ -b_p : p \notin B(w) \}$ if $\pi \neq \emptyset$, $\emptyset$ if $\pi = \emptyset$. The resulting schema is $\models^\mathcal{W}_w B\phi \rightarrow (\neg B\psi \rightarrow K-B\psi)$ for any $\phi, \psi$. 

Definition 5.14 there is a $q'$ such that $p' = -b_{q'}$ and $\neg \exists m \in M(q' \in m(w'))$ so that $w' \in -b_{q'}$. So by Definition 5.7, $m^{ni} \in M^I(w)$. Finally, by Definitions 5.14 and 5.10, $-b_p = \lbrack [\neg B \phi] \rbrack^M$. Since we have $m^{ni} \in M$, $m^{ni} \in M^I(w)$ and $-b_p \in m^{ni}(w)$, by Definitions 5.8 and 5.10, $\models_{w} K \neg B \phi$. □

In the proof of positive self-knowledge, we rely on fine-grained combined positive introspection methods $m^{pi} \circ m$ for each method $m$ the agent has. We could instead have used a simpler coarse introspection method $m^{pis}$ such that $m^{pis}(w, \pi) = \{ b_p : p \in B(w) \}$ for any $w, \pi$. Such a method is sufficient for positive self-knowledge (the proof is on the model of the one for negative self-knowledge). However, fine-grained introspection methods are necessary to derive knowledge of one’s knowledge, as we will see.

**Perfectly Confident Introspection: knowledge of one’s knowledge and partial knowledge of one’s ignorance**

**Definition 5.16.** A *confident introspector* model is a methods model for which there are positive and negative Confident Introspection methods $m^{pi}$, $m^{ni} \in M$ such that $m^{pi}(w, \pi) = \{ k_p : p \in \pi \}$ and $m^{ni}(w, \pi) = \{ -k_p : p \notin B(w) \}$.

For any $p$, if a method $m$ outputs the proposition $p$ ($p \in m(w)$), then $m^{pi} \circ m$ outputs the proposition that $k_p$ ($k_p \in (m^{pi} \circ m)(w)$). If no methods outputs $p$, then $m^{ni}$ outputs $-k_p$. The methods are introspection methods in which the outputs is the proposition that one knows (or fails to do so) instead of the proposition that one believes (or fails to do so).

**Theorem 5.11.** Confident Introspection. If $M$ is a confident introspector model, $\models_{w} B \phi \rightarrow BK \phi$ and $\models_{w} \neg B \phi \rightarrow B \neg K \phi$.

*Proof.* Evident from Definitions 5.10, 5.8 and 5.16. □

**Theorem 5.12.** Knowledge of one’s knowledge (4). If $M$ is a confident introspector model in a transitive frame, $\models_{w} K \phi \rightarrow KK \phi$ for any $\phi$.

*Proof.* We show that whenever the agent knows $p$ on the basis of some method $m$, she believes that she knows $p$ on the basis of $m^{pi} \circ m$, and we...
show that $m^{pc} \circ m$ is infallible if $m$ is. Since the agent knows $p$ on the basis of $m$, $m$ is infallible, and thus the agent knows that she knows $p$ on the basis of $m^{pc} \circ m$.

Suppose $\mathcal{M}, w$ are a confident introspecter model and a world such that $\models^\mathcal{M}_w K \phi$. By Definitions 5.10 and 5.8, there are $p, m$ such that $\llbracket \phi \rrbracket^\mathcal{M}_w = p$, $p \in m(w)$ and $m \in M^I(w)$. By Definition 5.16, there is a $m^{pc}$ such that $k_p \in (m^{pc} \circ m)(w)$. Moreover, suppose that $p' \in (m^{pc} \circ m)(w')$ for some $w'$ such that $wRw'$. By Definition 5.16, there is a $q'$ such that $p' = k_{q'}$ and $q' \in m(w')$. By the transitivity of $R$ and $m \in M^I(w)$, we also have $m \in M^I(w)$ (Lemma 5.1). Since $q' \in m(w')$ and $m \in M^I(w')$, $w' \in k_{q'} = p'$. So $(m^{pc} \circ m) \in M^I(w)$. Finally, by Definitions 5.14 and 5.10, $k_p = \llbracket Kp \rrbracket^\mathcal{M}_w$.

The result is the only one that requires a stronger assumption on the background accessibility relation than reflexivity. See section 5.2.3.

**Theorem 5.13.** Partial knowledge of one’s ignorance ($p5$). If $\mathcal{M}$ is a perfectly confident perfect introspecter, $\models^\mathcal{M} \neg B\phi \implies K \neg K \phi$ for any $\phi$.

Given subjectivity ($K \phi \implies B \phi$), the theorem is equivalent to a conditionalised version of axiom 5: $\models^\mathcal{M} \neg B \phi \implies (\neg K \phi \implies K \neg K \phi)$. I’m using “ignorance” here in a slightly unnatural way to refer to everything the subject fails to know. (Thus if $p$ is false $p$ is part of the subject’s “ignorance” in that sense.)

**Proof.** We show that whenever an agent fails to believe $p$, she believes that she does not know it on the basis of $m^{pc}$, and that $m^{pc}$ is infallible. The proof is analogous to the proof of negative introspection in Theorem 5.10; note that for any $w$, if $w \in \neg b_p$ then $w \in \neg k_p$ (Definition 5.14 and Theorem 5.2).

Partial knowledge of one’s ignorance ($p5$) is a very intuitive result. There has been much debate around axiom 5 of epistemic logic, according to which if one does not know $p$, one knows that one does not know it.
The intuition that has lead many to think that it was appropriate for knowledge is, I think, the following: ask an agent whether \( p \), she will “look up” her memory to see whether it contains \( p \), and if it does not, she will answer (rightly) that she does not know. But that is precisely the idea that our result cashes out: when a subject fails to know \( p \) because they fail even to believe \( p \), they know that they do not know \( p \).

Positive and Negative Confident introspection are introspection methods: they require an agent to be sensitive to his first-order beliefs. For instance, it would be more natural to define negative confidence as the combination of a psychological introspection method and a confidence method: whenever the agent does not believe \( p \), she believes that she does not, and whenever she believes that she does not, she believes that she does not know \( p \). We could introduce a negative confidence method \( m_{nc}^* \) such that \( n_{nc}^*(w, \pi) = \{ -k_p : -b_p \in \pi \} \). If the agent has a negative introspection method \( m_{ni} \), it is then easy to prove that \( m_{nc}^* \circ m_{ni} \) is infallible and to derive (p5). We additionally get: \( \models_{\mu} \BB \phi \rightarrow \BK \phi \) for all \( \phi \). This would make explicit the sense in which epistemic introspection depends on psychological introspection.

Unfortunately, we cannot implement the parallel idea for positive confidence within the limits of our simple models. Consider the method \( m_{pc}^* \) such that \( n_{pc}^*(w, \pi) = \{ k_p : b_p \in \pi \} \). (In intuitive terms: if the agent believes that she believes that \( p \), then she infers that she knows that \( p \).)

Suppose we have a methods model \( \mathcal{M} \) with two methods \( m, n \in M \) such that \( m \) outputs \( p \) and \( n \) outputs \( q \) at \( w_1 \) and neither outputs anything else at any other world: \( m(w_1) = \{ p \}, n(w_1) = \{ q \} \), and for any \( w \neq w_1 \), \( m(w) = n(w) = \emptyset \); suppose moreover that no other method of the agent outputs \( p \) or \( q \). The set of worlds in which \( p \) is believed is the same as the set of worlds in which \( q \) is believed, namely \( \{ w_1 \} \). We therefore have: \( m_{pi} \circ m = m_{pi} \circ n \). It follows that \( m_{pc}^* \circ m_{pi} \circ m = m_{pc}^* \circ m_{pi} \circ n \). This prevents us from proving that if \( m \) is infallible, \( m_{pc}^* \circ m_{pi} \circ m \) is infallible too, for \( n \) may be fallible. The difference between introspection from \( m \) (\( m_{pi} \circ m \)) and introspection from \( n \) (\( m_{pi} \circ n \)) is lost on the confidence method \( m_{pc}^* \), so to speak. One way to avoid the problem is construe confidence as directly
introspecting the lower-order method, as is done with \( m^{pc} \). Other ways require more refined models which we will not get into here.\(^{38}\)

However, we can partially capture the idea that confident introspection requires introspection in the following definition:

**Definition 5.17.** A normal confident introspecter model is a confident introspecter model \( \mathcal{M} \) such that \( \models_{\mathcal{M}} B\phi \rightarrow BB\phi \).

Suppose \( \mathcal{M} \) is a confident introspecter model that is not normal. We have a \( w \) such that \( \models_{\mathcal{M}} BK\phi \land \neg BB\phi \) for some \( \phi \) (Definition 5.17 and Theorem 5.11): the agents believes that she knows something without believing that she believes it. This is not only sub-ideal, but pathological.\(^{39}\) A natural class of normal confident introspecters is of course that of confident introspecters that are perfect introspectors.

### 5.5.5 Excellence

A third interesting class of models is that of excellent agents. An excellent agent is simply an agent whose methods are all infallible.

**Definition 5.18.** A excellent agent model is a methods model such that \( m \in M \rightarrow m \in M'(w) \) for any \( m, w \).

**Theorem 5.14.** Belief is knowledge (BK). If \( \mathcal{M} \) is an excellent agent model, \( \models_{\mathcal{M}} B\phi \leftrightarrow K\phi \).

**Proof.** The right-to-left direction follows from Subjectivity.

The left-to-right direction is evident from Definition 5.18, 5.8 and 5.10.

---

\(^{38}\) One way is to have fine-grained propositions. The outputs of methods are fine-grained (sentence-like) propositions. Roughly, we would have \( (m^p \circ m)(w_1) = \{Bp\} \) and \( (m^p \circ n)(w_1) = \{Bq\} \), so we have \( m^{pc} \circ m \neq m^{pc} \circ n \) as long as \( m(w) \neq n(w) \) for some \( w \). We can then define confidence as such that \( m^{pc}(w, \pi) = \{Kp : Bp \in \pi\} \) and derive knowledge of one’s knowledge. The other way is to characterise confidence as a constraint on agent’s methods, rather than as method itself. We say that for any introspection method \( m^{pi} \) the agent has, there is a confident introspection method \( m^{pc} \) defined as here. (See also the definition of normal confident introspecters below.) To model the idea would require meta-methods, i.e. methods to build methods.

\(^{39}\) Except if the agent has the concept of knowledge without having that of belief.
some $\phi$. There is a $p$ such that $[\phi]^n_M$ and $p \in B(w)$. Since $p \in B(w)$, there is a $m \in M$ such that $p \in m(w)$. By excellence, $m \in M_1(w)$. So $p \in K(w)$, and $\models^w K\phi$. □

For an excellent agent, believing is knowing, since all her beliefs are infallibly based. Correlatively, the only way such an agent fails to know something is by failing to believing it — while imperfect agents can fail to know something by having a false belief or by having a fallibly-based belief in it. That is the basis of the next result:

**Theorem 5.15.** Perfect knowledge of one’s ignorance (5). If $\mathfrak{M}$ is an excellent agent and a confident introspecter model, $\models^\mathfrak{M} K\phi \rightarrow K\neg K\phi$.

**Proof.** Suppose $\mathfrak{M}, w$ are a excellent perfectly confident introspecter model and a world such that $\models^w \neg K\phi$. By Theorem 5.14, $\models^w \neg B\phi$. By Theorem 5.13, $\models^w K\neg K\phi$. □

The result is again intuitive. However, it provides an illuminating perspective over the much-disputed axiom 5. While the axiom is assumed in many successful applications of epistemic logic, it faces a glaring and simple counter-example: false belief. If I mistakenly believe that my car keys are in my pocket, then I do not know that they are there, but (typically at least) I will not know that I do not know it. To the contrary, I (typically) think that I do know it. That does not reflect any irrationality on my part; nor is it plausible to say that I implicitly know that my keys are not there. Our result is in line with that idea: we derive knowledge of one’s ignorance for excellent agents, i.e. agents who cannot have false beliefs. At the same time, the result explains why (and when) it is safe to assume 5: namely, when the agent can be taken to be excellent with respect to the relevant facts. For instance, in many game-theoretic applications, it is assumed that the agents cannot have false beliefs about the game setup or draw false inferences. The simple S5 epistemic system is suited to that use.40

---

40. Revision note. Timothy Williamson pointed out another undesirable consequence of coarse propositions with respect to excellent agents. Since for excellent agents,
5.5. Results

5.5.6 Discussion

We have shown that under a natural set of idealisations, the axioms of a standard S5 epistemic logic hold. That ensures that many applications of epistemic logic can be recovered in methods models. Additionally, we have derived a number of principles linking knowledge and belief: Subjectivity, Self-knowledge, Partial Knowledge of One’s Ignorance and Belief is Knowledge.

But even though their stronger versions are equivalent to a single-operator S5 system, methods models provide an insight into the idealisations at work behind the S5 axioms. A widespread picture about epistemic logic is the following:

The axioms of standard epistemic logic represent an ideal of rationality, where rationality is a matter of internal or subjective coherence, as opposed to a matter of excellence, that is, objective success.

Our results suggest a radically different picture.

We have three sets of derivations that are independent: (a) Perfect Reasoning, (b) Positive epistemic introspection and partial negative epistemic introspection (4 and p5), (c) Excellence. Negative epistemic introspection is obtained from a combination of (b) and (c). But it is important to note that the results in (b) and (c) do not assume perfect reasoning, nor does (c) assume introspection or confidence. We thus have three distinct sets of idealisations at play.

Moreover, it can be shown that while the Perfect Reasoning methods are non-informative, Introspection and Confident Introspection are informative, in the sense that they “narrow down” the sets of possible worlds compatible with what the agent believes and knows. (That is, if a possible world is incompatible with an agent’s belief based on $m^D \circ m$, that possible world is also incompatible with some belief of the agent based on believing is knowing, $B_\phi$ and $K_\phi$ express the same proposition for any $\phi$. This precludes the possibility of a modest excellent agent that knows that $\phi$, believe that they believe that $\phi$, but do not believe that they know that $\phi$, for by the equivalence of $B_\phi$ and $K_\phi$, $BB_\phi$ entails $B_\phi$. \footnote{...}
on $m$; the same does not hold for $m^{pi} \circ m$.) See Appendix C.5 for a formal
definition of the relevant notions of information and informativeness.

Together, these remarks suggest the following picture:

1. Deduction and Reason (axiom $K$) are properly a matter of pure
rationality or internal coherence. They do not provide information,
but make explicit the information the subject has.

2. Introspection and Confident Introspection (axioms $4$ and $p5$) are a
matter of excellence with respect to the inner. Both assume that the
agent has reliable ways to find out about its own internal states.
Both are information-purveying methods. Thus satisfying axiom $4$
or $KK$ is not simply a matter of internal coherence.

3. Excellence (axiom $5$) typically requires excellence with respect to
the outer. Extreme cases aside, it requires one’s methods to provide
information. \footnote{The extreme case is that of an agent who has only Reason and Deduction.} It is not a simple matter of internal coherence either.

The methods approach thus exhibits a distinction between three group of
idealisations behind standard epistemic logic: pure rationality, internal
excellence and external excellence. They can thus provide a guide as to
when it is appropriate to assume the axioms to hold.

5.6 Conclusion

Methods models provide a formal representation of knowledge that
rests on the methods-infallibilist conception of knowledge. The conception is in line with various trends in mainstream epistemology, such as
safety theories or some variants of virtue epistemology. I have argued
that it is suited to model and think about classical epistemological is-

dues such as the Gettier problem or inductive knowledge. Because the
intuitive notion of method, or basis of belief, takes a centre stage in the
models, they should prove more amenable to epistemologists than the
standard Hintikka models. However, I have also shown that standard
epistemic logic systems can be recovered from methods models with a
5.6. Conclusion

A series of natural idealisations of agents and a constraint on possibility. The models thus provide a new vindication of the standard axioms and an illuminating perspective on why and when they hold or not. They should consequently prove useful to formal epistemologists as well.

The models can be further developed in a range of directions, and much needs yet to be done. On the formal side, they should be studied syntactically, starting from the algebra we sketched and by introducing an operator to express infallibility. Soundness and completeness properties should be established. Relatedly, the models may be usable as models for the Logic of Proofs. A further important formal development is to build variants of the models that integrate common treatments of referential opacity, which will most likely require us to leave the ground of neighbourhood semantics.

On the epistemology side, four developments can be mentioned. First, our methods are only methods of belief formation. Methods of belief inhibition and revision should also be considered. The former would allow holistic constraints on the belief system (e.g., if a method produces a belief that $p$ and another a belief that $\neg p$, both beliefs are suspended); but for that reason, they may be hard to accommodate formally. The latter would require us to recast our models in dynamic terms, with temporal slices of agents being characterised by the set of beliefs reached at each point or by distinct sets of methods. Second, we have only considered fine-grained Introspection and maximal Confidence. Variants of Introspection that fail to discriminate between beliefs produced by a range of similar methods should be discussed, as well as cautious agents whose epistemic confidence extends only to the beliefs produced by a subset of their methods. Third, our methods can straightforwardly be used to model conditional belief and hypothetical reasoning. If an agent has a method $m$ such that $p \in m_w(\pi)$, then (at $w$) she conditionally believes that $p$ on the hypothesis that $\pi$. However, the Introspection methods as defined here are unsuitable for that purpose (they imply that, for any $p$, the agent conditionally believes that she believes that $p$ on the hypothesis that $p$). Accordingly, we may want to redefine them as a class
of non-inferential methods. Fourth, the idea of a reliability measure over methods that we have sketched in section 5.3.3 should be investigated in order to see whether it can yield an interesting notion of epistemic probability.

These developments only concern the single-agent case. A further one is of course to study multi-agent settings and to characterise common knowledge in method terms. Mind-reading methods may prove relevant here, as well as perspectives (section 5.2.1).

Finally, the methods approach need not be restricted to epistemology. Alongside methods of belief formation, one may characterise an agent by a set of methods for decision, whose inputs are a set of premises (and perhaps a set of aims) and whose outputs are actions. Truth is here replaced by success. In the epistemological case, the approach induces a shift of focus from individual beliefs to classes of beliefs formed in the same way. In the practical case, we get an analogous shift of focus from particular intentions or actions to classes of actions that result from the same policy. The latter kind of focus is already familiar from rule utilitarianism and virtue theories. Methods models may provide useful representations of such ideas.
Part IV

Semantics
In this part of the thesis I want to examine how method infallibilism sheds light on the ordinary conception of knowledge. I take the ordinary conception of language to be manifested in knowledge claims, that is, the use of “know” in natural languages. As I have stressed (1 and 4.3.2), I do not think that the notion of method, in the sense used here, is part of our ordinary conception. But I think that impossibility, in the sense in which knowledge requires impossibility of error, is part of our ordinary conception. The hypothesis I will put forward is that the link between infallibility and knowledge is reflected in ordinary language through links between knowledge claims and possibility claims.

In recent literature, the idea that considerations of possibilities of error and knowledge ascriptions are related has been particularly investigated by contextualists about knowledge. I will use a discussion of major contextualist views as a springboard to introduce my own views. In the present chapter I introduce contextualist theories and raise a number of difficulties for them. In the next one I formulate a semantics based on the method infallibility idea. The semantics can be construed in a contextualist way, and if so avoids a number of problems that beset existing brands of contextualism. However, it can also be construed in an invariantist way, and I will remain ultimately neutral on the issue.

Section 6.1 introduces Epistemic Contextualism in general, and iso-
lates Modal Epistemic Contextualism as our topic of interest. Sections 6.2 and 6.3 present Lewis and DeRose’s variants of modal epistemic contextualism, respectively. We discuss Lewis’s notion of attending to possibilities in some detail. We emphasise that both Lewis and DeRose postulate rules of context-sensitivity that are specific to “know”. In DeRose’s case, we also argue that the rule has no analogue in language, by drawing analogies with gradable adjectives. Sections 6.4—6.6 raise three problems for Lewis’s and DeRose’s contextualism. The first is, again, the necessary truths problem (see sec. 4.1.1): both views are proposition-centred, and consequently run into trouble with necessary truths. I object to Lewis’s way of dodging the problem and to Blome-Tillmann’s alternative proposal for avoiding it and conclude that methods-based contextualism is preferable. The second problem is that the shifts in intuitions that motivate contextualist views do not occur when we merely attend to error possibilities, as Lewis thought, but when we take them seriously. Drawing on Blome-Tillmann’s idea that taking a possibility seriously is not presupposing it not to obtain, I consider the hypothesis that taking a possibility seriously is simply not believing it not to obtain. But the proposal — as well as Blome-Tillman’s one — fails in view of third-person sceptical arguments. An adequate account of taking seriously is still lacking. The third problem arises from consideration of subjects and cases that are not mentioned in a conversation. They show that Lewis’s account and Blome-Tillmann’s variant of it violate the factivity of knowledge.

Some main objections to epistemic contextualism are not discussed in this chapter, notably the claim that it is irrelevant to epistemology (Sosa, 2000b), and the claim that it entails implausible semantic blindness (Schiffer, 1996; Hawthorne, 2004, 98–111; Williamson, 2005a). Rather, I will discuss them in the light of the alternative contextualist proposal formulated in the next chapter.
6.1 Epistemic contextualism

Epistemic Contextualism is the view that “knows” makes different contributions to the utterances in which it occurs in relevantly different contexts.\(^1\) If the variation is not vacuous, there is a case \(\alpha\) (involving a particular subject, world, time, and proposition) such that in some context it is true to say that the subject “knows” the proposition, and false to say it in another. In our terminology of cases and conditions (sec. 3.1), epistemic contextualism is the view that “knows” expresses different conditions in relevantly different contexts. Invariantism, by contrast, is the view that “one knows \(p\)” expresses a unique condition at all contexts.\(^2\)

(Throughout the chapter I follow Blome-Tillmann’s (2009b) practice of using hybrid quotation to avoid both cumbersome metalinguistic formulations and misleading or literally false claims. When I say that it is true to say, in context \(c\), that the subject “knows” the proposition, I mean that some sentence “\(S\) knows \(p\)” is true in the context, where \(S\) names of the subject and \(p\) expresses the proposition. Thus on its hybrid quotation use, ““knows”” takes the semantic value that “knows” takes in the context under discussion.\(^3\) Double quotes are often used as corner quotes.)

---


\(^2\) Recall that our conditions are on proposition-centred cases (sec. 3.1). Thus “Peter knows that \(p\)” uttered at the context \(C\) ascribes a condition \(K_C\) to the case \(\alpha = < w_C, t_C, p >\) where \(w_C\) and \(t_C\) are the world and time of \(C\), respectively. \(K_C\) is the condition on proposition-expressed cases that the English “knows” expresses in \(C\). Invariantism is the view that \(K_C = K_{C'}\) for any \(C, C'\).

Invariantists can hold instead that “knows” expresses a unique set of conditions at all contexts, if “know” is partly indeterminate. They may harmlessly grant that in some peculiar contexts (e.g. some computer science papers or cases of ironic deference) “know” is used expresses some other condition. The terminology “contextualism”/“invariantism” is due to Unger (1984).

\(^3\) See Davidson (1983), Cappelen and Lepore (1997) (who introduced the label)
Contextualism is motivated by the apparent existence of such pairs of cases. Three broad types of pairs have been put forward: sceptical scenarios (6.1 a–6.1 b), ordinary cases (6.2 a–6.2 b), and ordinary cases in which “lottery inferences” are drawn (6.3 a–6.3 b): 4

(6.1) (a) Sally knows that the book is red. [said in an ordinary context.]

(b) Sally does not know whether the book is red. [said in the midst of a convincing discussion of external world scepticism]

(Cohen, 1988, DeRose, 1995)

(6.2) (a) Keith knows that the bank is open on Saturday. [said in an ordinary context.]

(b) Keith does not know whether the bank is open on Saturday. [said in a context in which it has been made salient that the bank could change its hours.]

(DeRose, 1992, 913; see also Cohen’s 1999, 58 Airport case)

(6.3) (a) John knows that he will in Oxford next month. [said in an ordinary context]

(b) John does not know whether he will be in Oxford next month. [said in a context in which it has been argued that John does not know that he will not be one of the few unlucky people who suffer an unexpected heart attack.]

(variant of Vogel, 1990, 20–1)

(In each pair the situation with respect to which knowledge is attributed is meant to be the same. The situation and the ordinary contexts are filled in so that the attribution is one that we would naturally accept. The third group of cases can easily be seen as a variant of the second in which the sceptical intuition is supported by a use of deductive closure. The first may be seen as an extreme version of the second or the third.)


4. I call “lottery inference” a deductive inference whose conclusion is what Hawthorne (2004, 5) calls a “lottery proposition”.
6.1. Epistemic contextualism

Say that an application of “knows” is paradigmatic if there is a wide and robust disposition to apply the concept in cases of its kind, and there is a wide and robust agreement that the concept is truly applied in that case. (And similarly for denials of knowledge.) To be precise, one should distinguish the case in which “know” is applied from the case to which “know” its applied. What we call paradigmatic cases of application here are cases in which “know” is used.

Epistemic contextualists typically take pairs such as (6.2 a−6.2 b) to be part of the paradigmatic cases of application of “know” (DeRose, 1995, 172).5 (Pairs (6.1 a−6.1 b) are certainly not part of the paradigm cases: though sceptical inclinations are widespread, they are not at all universally shared.) They argue that our judgements in such cases are right, and conclude that contextualism is true.6

Let us call $K$−conditions the conditions expressed by “knows” in different contexts. We write $K_C$ the condition expressed by “know” in a context $C$. (On the invariantist view, there is only one $K$-condition: $K_C = K_{C'}$ for any $C, C'$..) Contextualists have different views on the dimensions along which $K$−conditions vary. Some provide justificationist (Cohen, 1999, 59) or explanationist accounts (Rieber, 1998, 194–8) of the variety of $K$−dimensions. Here I am interested in contextualist accounts according to which the $K$−conditions vary in considering different ranges of possibility of error (DeRose, 1995, Lewis, 1996, 553–4). Call them Modal Epistemic Contextualist Accounts. The core idea of modal epistemic contextualism is that a conversational context determines a set of relevant alternative cases, so that it is true to say in that context that one “knows” something only if error (of a certain type) is avoided in those cases.

In the following sections we present Lewis’s (1996) and DeRose’s (1995; 2009) brands of modal epistemic contextualism in some detail. But

5. Buckwalter’s (forth.) experimental results challenge this view.
6. Let $\alpha$ be the target case. Let $\beta$ be a case in which $S_\alpha$ is said to “know” that $p_\alpha$ in $\alpha$, and let $\gamma$ be a case in which $S_\alpha$ is said to “know” that $p_\alpha$ in $\alpha$. We assume that both utterances are true. Suppose for reductio that utterances of “know” in $\beta$ and $\gamma$ express the same condition $K$. Since the claims in $\beta$ and $\gamma$ are true, we have $Ka \land \neg Ka$, which is impossible. Thus utterances of “know” in $\beta$ and $\gamma$ express different conditions.
let me summarise the important points here. Their accounts share three
significant features. (1) They take “know” to be context-sensitive in virtue
of rules that are specific to it and perhaps to other epistemic terms. (This
may seem trivial, but it is not. In the account we spell out in ch. 5, “know”
is context-sensitive in virtue of rules that govern the context-sensitivity
of counterfactuals and modals as well.) Lewis’s rules for “know” are
analogous to those for contextual domain restriction. But DeRose’s rule
for “know” does not seem to have any analogue in language. (2) They
are proposition-centred: one knows $p$ only if one avoids errors about $p$.
(3) They are method-relative. On Lewis’s view, only errors that $p$ based on
the same experience are relevant. On DeRose’s view, only errors based
on similar methods are relevant. They differ on two important counts.
(4) On Lewis’s view, context-shifts are automatic: when possible cases are
brought to attention, they are thereby made relevant. On DeRose’s view,
context-shifts can be resisted. (5) On Lewis’s view, the context determines
a set of relevant alternative cases in which error should be avoided. On
DeRose’s view, it determines a strength of epistemic position, which in
turns determines a set of relevant cases for any putative knowledge case.

6.2 Modal Epistemic Contextualism: Lewis

6.2.1 Outline

Lewis’s (1996, 553) states his account is as follows:

S knows that $P$ iff S’s evidence eliminates every possibility in
which not-$P$ - Psst! - except for those possibilities that we are
properly ignoring.

“We” are the ascribers in some context. Lewis “bends the rules” (566) of
use and mention. We should read:

S “knows” that $p$ in $w$ in $C$ iff S’s evidence in $w$ eliminates every
not-$p$ possibility except the ones that are properly ignored in $C$. 
In Lewis’s use, my current “evidence” “eliminates” a given possibility iff in that possibility, I do not have the experiences and memories I currently have (553). Possibilities are sets of centred worlds (552, more on this shortly). Finally, the context restricts the quantifiers “every” to a set of relevant possibilities, the ones that are “not properly ignored” (553).

Two idiosyncratic features of Lewis’s account are that he does not require belief (556) and that his propositions are coarsely individuated as sets of possible worlds, so that if one knows a necessary truth one knows all of them, for instance (552). Both ideas are unpalatable, but I will not discuss them here.

Lewis gives eight rules that indicate how a context determines a set of relevant cases in which error must be avoided. Here are the rules, adjusted to our terminology of cases (sec. 3.1). Four rules are prohibitive: they force some possibilities into consideration. Four others are defeasibly permissive: they defeasibly allow speakers to ignore some possibilities. Suppose that “knowledge” is ascribed with respect to a case $\alpha$ (or denied with respect to a case $\alpha$) in a context $C$:  

**Rule of Attention**  A possibility attended to in $C$ cannot be ignored. ([Lewis, 1996], 559)

**Rule of Actuality**  $\alpha$ cannot be ignored. ([Lewis, 1996], 554)

**Rule of Belief**  if in $\alpha$, $S_\alpha$ believes or ought to believe a possibility $p$ to obtain, then $p$ cannot be ignored. ([Lewis, 1996], 555)

**Rule of Resemblance**  if $p$ resembles $q$ in a way that is salient in $C$, and if $q$ cannot be ignored in $C$ (in virtue of other rules than the rule of resemblance), then $p$ cannot be ignored either. ([Lewis, 1996], 556)

**Rule of Reliability**  Possibilities in which reliable faculties fail may be ignored. (Faculties that are reliable in $C$? in $\alpha$? in Lewis’s and his readers’s actuality?) ([Lewis, 1996], 558)

**Rule of Method (1)**  Possibilities in which a sample is not representative may be ignored. ([Lewis, 1996], 558)

**Rule of Method (2)**  Possibilities in which the best explanation of the evidence is not true may be ignored. ([Lewis, 1996], 558)
Rule of Conservatism Possibilities that are commonly ignored in the community of C may be ignored. (Lewis, 1996, 559)

Lewis uses an ambiguous “we” that can refer either to speaker or subject of knowledge ascription. When Lewis writes, in the second Rule of Method, that speakers can ignore possibilities in which the best explanation of “our” evidence is not the true explanation, I take him to refer to the ascribee’s evidence. But when he says, in the Rule of Conservatism, that speakers can ignore possibilities that are commonly ignored in “our” community, I take him to refer to the speaker’s community. I have not been able to disambiguate the rule of reliability. My best guess is that he means reliability in the subject’s case. 7

6.2.2 Subject-sensitive vs. context-sensitive rules

Cohen (1998, 293) usefully distinguishes two kinds of rules in Lewis’s collection. Some rules are subject-sensitive: they count cases in or out in virtue of features of the subject (Actuality, Belief, and perhaps Reliability and Methods). Others are speaker-sensitive: they count cases in or out in virtue of features of the speaker’s context (Attention, Resemblance, Conservatism).

Cohen (1998, 297) objects to the Rule of Resemblance being speaker-sensitive, for the rule is meant to deal with Gettier cases, and whether a subject in a Gettier case “knows” is not a context-sensitive matter. (That is, if α is a Gettier case, there is no context C such that K_Cα.) He illustrates the point as follows. A man sees a sheep-shaped rock and comes to believe that there is a sheep in the field; as luck has it, there is a sheep behind the rock.

7. Discussing whether the Rule of Reliability is an instance of the Rule of Conservatism, Lewis (1996, 559) writes: “the only extra work done by the Rule of Reliability would be to cover less familiar [. . . ] reliable processes, such as processes that relied on extrasensory faculties.” Maybe what Lewis has in mind is this. If we (him and his readers) didn’t ignore occasional malfunction of ordinary faculties like vision, no ordinary “knowledge” ascriptions could come out true. Similarly, when considering a fictional story in which somebody has extrasensory faculties, we can truly ascribe “knowledge” to the fictional character, even though this requires ignoring occasional glitches in the character’s fictional world. If that is the idea, then he means reliability in the subject’s case.
rock (Chisholm, 1977, 105). On Lewis’s view, it is false for us to say that the man “knows” that there is a sheep, because his situation resembles saliently to us to one in which he has the same experience but there is no sheep. Now suppose that somebody next to the man ascribes him “knowledge”. To that ascriber (as opposed to us), it is not salient that the man’s situation resembles a situation in which he has the same experience but there is no sheep. Lewis has to say that the “knowledge” ascription is true, which seems clearly wrong (Cohen, 1998, 297). There are two replies for Lewis here. One is to deny that saliently resembling requires that a resemblance be salient to the speaker. Lewis could argue that what makes a case saliently resemble another is that some respect of resemblance is salient to the speaker, even though the resembling cases themselves are not salient to her. The relevant respect in the sheep situation is that the man is forming his belief on the basis of seeing that object [thinking of the rock] in the field. Unbeknown to the speaker, the no-sheep situation resembles the actual one in that respect, which is why her attribution comes out false even though she does not realise it. An alternative reply is to split the Rule of Resemblance into a subject-sensitive rule and a speaker-sensitive one. By the first rule, a case that is close to the subject’s case is not properly ignored, as in safety conditions. By the second, a case that saliently resembles the subject’s case is not properly ignored. The latter solution is equivalent to keeping the original Rule of Resemblance and stipulating that some minimal respects of resemblance are salient in every context.

Cohen’s distinction between speaker-sensitive and subject-sensitive rules easily leads to the impression that subject-sensitive rules are context-independent rules. It is worth stressing that they are not. “The subject” in subject-sensitive rule is the subject under discussion. Thus what cases

---

8. Note besides that if, as Cohen assumes, saliently resembling entailed that the speaker is aware of the saliently resembling cases, the Rule of Attention would make the Rule of Resemblance redundant.

9. More precisely, a case that saliently resembles the subject’s case or another case that is not properly ignored in virtue of rules other than the Rule of Resemblance. (A similar qualification may or may not be introduced in the closeness rule.)
are made relevant by these rules crucially depends on a feature of the context, namely, what subject is being discussed. As we will see, this is the reason why factivity fails on Lewis’s account (6.6).

6.2.3 Lewis’s rules are specific to “know”

Lewis’s rules are specific to “know”. They determine a set of relevant alternative cases whose only purpose is to evaluate “knowledge” claims, and perhaps claims involving related epistemic terms, such as epistemic modals, “prove”, “see” and so on. At one point Lewis (1996, 553) suggests a different account, on which the relevant cases are just those not properly presupposed not to obtain, where the relevant notion of presupposition is Stalnaker’s (1974/1999). On this view, the set of not properly ignored cases would be determined by the general mechanism of pragmatic presupposition. But that cannot be the account Lewis has in mind (as Schaffer, 2004a, 99n notes). Suppose Alice asks Bob:

(6.4) Does your wife know that you are here?

On any view, in uttering (6.4), Alice presupposes that Bob is here. And at least as far as the linguistic notion of presupposition goes, it is proper for Alice to do so: Bob is here and it is commonly “known” between them that he is — for some relevant value of “know”. If presupposition is the guide to proper ignoring, all cases in which Bob is elsewhere are properly ignored. In all the remaining cases, Bob is here. It follows that in all the remaining cases in which Bob’s wife has the evidence she actually has, the proposition that Bob is here is true. So by Lewis’s account, it is true in Alice’s context to say that Bob’s wife “knows” that Bob is here, whatever Bob’s wife evidence is. More generally, speakers would not be able to

10. Cf. Lewis (1996, 553): “We can restate the definition. Say that we presuppose proposition Q iff we ignore all possibilities in which not-Q. To close the circle: we ignore just those possibilities that falsify our presuppositions. Proper presupposition corresponds, of course, to proper ignoring. Then S knows that P iff S’s evidence eliminates every possibility in which not- P -Pst! - except for those possibilities that conflict with our proper presuppositions.” Lewis refers in a footnote to Stalnaker’s (1973; 1974/1999) and to his own (1979/1983).
state truly that someone fails to “know” \( p \), for any \( p \) that the speakers themselves commonly “know” to be true. A crazy result.

Lewis’s account is in effect incompatible with the presupposition account of proper ignoring. Consider for instance his treatment of Lehrer’s Nogot case (Lewis, 1996, 557). Smith believes that someone in his office owns a Ford because Nogot pretends to have one; Nogot does not, but Havít is also in the office and owns one. Lewis says that Smith does not know, because a possibility in which Smith has the evidence he actually has but Havít does not have a Ford either saliently resembles actuality, and thus cannot be ignored. But of course we presuppose (in the very description of the case) that Havít has a car, so we presuppose the possibility not to obtain. Yet we cannot properly ignore it, by Lewis’s rules. So Lewis’s rules are not tied to speaker’s presuppositions; they are specific to “know”.  

6.2.4 Possibilities and attending to possibilities

Lewis’s notion of possibilities raises thorny issues (Hawthorne, 2004, 63–4, Douven, 2005). They need not be maximally specific (as centred worlds and our cases are), but they need to be specific enough for what

---

11. Can Lewis’s suggestion be rescued by considering other mechanisms that presupposition? Lewis refers to his (1979/1983) in addition to Stalnaker’s work. Lewis’s notion of “conversational score” (1979/1983, 238–40) is wider than Stalnaker’s notion of presupposition: it includes the presuppositions of participants, but also elements representing the salient domains of quantification, points of spatial and temporal reference, standards of precision, and so on. So we can say that the “knowledge”-relevant set of cases is part of the conversational score. But that does not make it less specific to “know”: the relevant set of cases is just the “knowledge”-specific component of the conversational score.

The discussion in Lewis (1979/1983, 246–7) may suggest that there is a unique “possibility” component of the conversational score that is relevant for modals in general in addition to “knowledge”. Could we thus say that the relevant set of cases is just the possibility component of the score, and thus not specific to “know”? No, because we need a variety of possibility sets to account for cases in which various flavours of modalities are mixed in a same breath, as in “She must make sure that they cannot forget.” Once we have different “possibility”-sets for different modals, the one relevant to “knowledge” may be specific to “know”. This is what Lewis’s set of specific rules suggest.

One option left is that the set relevant to “knowledge” is also the one relevant to other expressions. This is the one I defend in ch. 7.
he says about them: “a possibility will be specific enough if it cannot be split into sub-cases in such a way that [. . .] anything we are going to say before we are done applies to some sub-cases but not to others” (Lewis, 1996, 552).

It is not clear that possibilities can be both specific enough and unspecific enough to be compatible with everything he says about them. As Hawthorne (2002, 243) notes, possibilities cannot be maximally specific, otherwise we could not attend to them and the rule of Attention would have no application. So they need to be unspecific enough that we attend to them. But for ordinary “knowledge” ascriptions to come out true we need them to be specific enough to exclude cases of error based on the same experience. Consider:

(6.5) Davide. — The keys are not in the room. I have searched it.
Camila. — Are you sure they are not in one of the desk drawers?
D. — Yes, I have checked the three drawers. I know that the keys are not in the room.

The desk has only three drawers, Davide has checked them, the keys are not in the room, and everything else is as normal as can be. This is a case in which Lewis wants to say that Davide’s last utterance is true. Now on the face of it, the possibility that Camila raises (and attends to) is the possibility that the key is in one of the desk drawers. But the possibility has many sub-cases in which Davide would have the same experience as he actually has:

There is a fourth hidden drawer in the desk and the key is there.
The key is there but becomes invisible when one looks at it.
The key is momentarily microscopic.

And so on. If that is the possibility Camila is attending to, the sub-cases cannot be ignored (by the Rule of Attention), and Davide’s utterance comes out false.12 By contrast, Davide’s utterance comes out true if Camila only brings in the possibility that the key is in one of the desk drawers,

12. Can we not say that the problematic sub-cases are ruled out by the Rules of Reliability, Conservatism, and Methods? That would mean that permissive rules can
but without being a secret hidden one, microscopic, and so on. But it is unclear whether Camila is attending to anything as specific as that.

One solution is to amend the Rule of Attention as follows: if a possibility \( p \) is attended to in \( C \), then at least some sub-case of \( p \) should be considered. But that is both incomplete (which sub-case?) and too generous. Suppose Davide has only checked one drawer; if we only consider the sub-case in which the key is in that drawer, his utterance will still come out true, contrary to intuition. A better amendment is to say that close sub-cases of \( p \) cannot be properly ignored. This would bring Lewis’s Rule of Attention close to DeRose’s Rule of Sensitivity (to which we return below). A third option is to claim that Camila did have something more specific in mind. What she was attending to was something like the possibility that the key was in one of the drawers in a normal or unsurprising way. She need not explicitly think of the possibility as being a normal one, nor does she need to have in mind specific ways in which the possibility would be a normal one. The mere fact that she is not disposed to think of such sub-cases as the ones we mentioned, and that she would be surprised if they were mentioned, is sufficient to make it the case that she is attending to a more specific possibility in which the key is in one of the drawers. The second and third options need not be seen as competitors. It may be said that when one thinks of \( p \), one typically attends to close possibilities that \( p \).

Douven (2005) has raised a related objection to Lewis. On Lewis’s view, overt claims of fallibilism such as (6.6) “sound wrong” (1996, 550), and his semantics is designed for them to come out false.

(6.6) He knows that \( p \), yet he has not eliminated all not-\( p \) possibilities.

On Douven’s view (574), Lewis’s semantics is supposed to secure the result as follows: by mentioning not-\( p \) possibilities, the second part of the sentence brings them to salience, and the first part comes out false. But, if the speaker has no specific not-\( p \) possibility in mind, she may fail to defeat the Rule of Attention. But if Lewis allowed that, the same could be said about sceptical possibilities, so it is unclear that his explanation of the power of sceptical arguments would stand. A refined version of this reply would be: permissive rules can allow ignoring some sub-cases of a possibility attended to; they cannot allow ignorance of all of its sub-cases. I do not explore this line of reply here.
bring one to salience (575). So Douven argues that (6.6) could come out true, provided there are uneliminated not-\(p\) possibilities that the subject has not eliminated (575).\(^{13}\)

Lewis may reply in three ways here. One is concession: he could accept that it is possible to state overt fallibilism as long as one does not bring specific possibilities of error into focus.\(^{14}\) A second is to say that the mention of not-\(p\) forces one to attend to close not-\(p\) possibilities. A third is to deny that (6.6) needs to come out false in the way Douven presupposes. Lewis’s idea is that quantification over possibilities is restricted (553). Thus when a speaker utters (6.6), the second part of her utterance is true iff there are not-\(p\) possibilities in the domain of her context. Douven is right to say that there might be uneliminated not-\(p\) possibilities (in the domain of our context), but it does not follow that there are not-\(p\) possibilities in the domain of the speaker’s context. So if the speaker has no specific not-\(p\) possibility in mind and the second part of the utterance fails to bring any to the domain of the context, then the second part of the utterance comes out false as well.\(^{15}\)

### 6.2.5 Formal version of Lewis’s view

It will be useful to restate Lewis’s semantics in a more formal way. Recall that in our framework, cases \(\alpha, \beta, \ldots\), are centred worlds with a

\(^{13}\) Douven suggests a possible reply on behalf of Lewis: if a speaker entertains the proposition that there are uneliminated not-\(p\) possibilities, then possibilities that would make it true that there are uneliminated possibilities would thereby be indirectly attended to (576). But not only would the suggestion not match with linguistic practice (i.e., with the fact that sceptics do need to bring quite specific error possibilities to attention), as Douven argues (578), it would also lead to sceptical results that Lewis is not prepared to accept. In (6.5), it would imply that Davide’s last utterance is false, since all the bizarre variants of Camila’s envisaged possibility would be indirectly attended to.

\(^{14}\) The reply adapts Lewis’s (1996, 562) idea that he can claim *in general* that one “knows” by ignoring without being able to claim any *particular* instance of this.

\(^{15}\) The third reply may be coupled with the second as follows: either the mention of not-\(p\) succeeds in bringing close not-\(p\) cases to mind or not. If it does not, the relevant set is not expanded, and the second part of the utterance comes out false — with details of the case appropriately filled in. If it does, and if some of the cases are error ones, the first part comes out false. In both cases the utterance is false.
proposition singled out (section 3.1.1). \( w_\alpha \) is the world of \( \alpha \), \( S_\alpha \) is the subject of \( \alpha \), \( p_\alpha \) is the proposition of \( \alpha \). We write:

\[ T_\alpha \ p_\alpha \] \( \) is true in \( \alpha \).

\[ H_\alpha \ S_\alpha \] has experience (and memories) \( e \) in \( \alpha \).

\[ E_\alpha \ S_\alpha \] has experience \( e \) in \( \alpha \) while \( p_\alpha \) is false: \( H_\alpha \land \neg T_\alpha \). (This is “error” in a quite extended sense, since the proposition \( p_\alpha \) need not be believed in \( \alpha \).)

\[ K_C \] the \( K \)–condition expressed by “knows” in \( C \).

\[ P_C \] the relevant alternative cases in \( C \), that is, the union of the possibilities that are not properly ignored in \( C \). \(^{16}\)

**Lewis’s Modal Contextualism** For all \( C, \alpha \), \( K_C \alpha \) iff for some \( e \): \( H_\alpha \) and for all \( \beta \in P_C \), if \( p_\beta = p_\alpha \) then \( \neg E_\alpha \beta \).

For all \( C, \alpha \), \( K_C \alpha \) iff \( \exists e (H_\alpha \land \forall \beta ((\beta \in P_C \land p_\beta = p_\alpha) \rightarrow \neg E_\beta)) \).

# 6.3 Modal Epistemic Contextualism: DeRose

## 6.3.1 Outline

On DeRose’s (1995, 29) view, a conversational context affects the strength of the epistemic position one needs to be in in order to satisfy “know” in that context. DeRose’s official characterisation of strength of epistemic position is only through intuitions about “knowledge” claim (1995, 30; 2009, 7–8): for any case \( \alpha, \beta \), if the comparative conditional “if \( \alpha \) is a case of knowledge, \( \beta \) is one too” is true, then the subject’s epistemic condition in \( \alpha \) is at least as strong as the one in \( \beta \). (The conditional is assumed to be context-insensitive, since comparative epistemic strength is not affected by context.) But DeRose proposes a “partial and rough and ready” (1995, 34) characterisation of strength of epistemic position in modal terms:

An important component of being in a strong epistemic position with respect to \( p \) is to have a belief as to whether \( p \) is

\(^{16}\) We call them the *privileged cases* of \( C \), sec. 6.6.3 below.
true match the fact of the matter as to whether \( p \) is true, not only in the actual world, but also at the worlds sufficiently close to the actual world. That is, one’s belief should not only be true, but also should be non-accidentally true, where this requires one’s belief as to whether \( p \) is true to match the fact of the matter at nearby worlds. The further away one gets from the actual world, while still having it be the case that one’s belief matches the fact at worlds that far away and closer, the stronger a position one is in with respect to \( p \).

We can picture the idea schematically as follows (Blome-Tillmann, 2009a, 384). Assume that the similarity between worlds determines a distance measure. The strength of a subject’s epistemic position in a belief case is given by how distant from the world of the case are the closest worlds in which the subject’s belief fails to match the fact.\(^{17}\) The strength of one’s epistemic position in a case is not a context-sensitive matter. What a context selects is a threshold such that for a case to satisfy “know” in that context, the subject’s epistemic position in the case must be above threshold. The threshold can be thought of as a particular distance. Match between belief and fact require both avoidance of error and avoidance of ignorance: the subject believes \( p \) if and only if \( p \).\(^{18}\)

### 6.3.2 Closeness and methods

DeRose’s similarity relation must be significantly tailored for the rough and ready characterisation to work reasonably. The adherence condition easily leads to excessive denials of knowledge (see also 3.4.5). Suppose I glimpse a bird by the window: I form the true belief that a bird

\(^{17}\) By assumption, the strength of one’s epistemic position in non-belief cases is null. The account does not straightforwardly apply to \textit{de se} propositions (see 3.1.1). We ignore them when discussing DeRose’s view.

\(^{18}\) Presumably time is more or less fixed: error and ignorance should be avoided at close worlds at the time of the case, not wildly earlier or later. My childhood beliefs in the existence of Gnomes at close worlds do not lessen my epistemic position with respect to their inexistence now. Alternatively, DeRose’s closeness could be conceived as a relation between cases, instead of a relation between worlds.
6.3. Modal Epistemic Contextualism: DeRose

passed by. But there are very close worlds in which the bird passed by slightly earlier or later and I did not notice it. In those close worlds, I do not avoid ignorance. DeRose would thus have to say that my epistemic position is very weak in such case. Similarly for the avoidance of error condition, if there are close cases in which I would falsely believe that there is a bird but on the basis of a different method (say, hearing a whistle). To avoid both problems, DeRose’s closeness relation must give a heavy weight to similarity in methods formation (1995, 20–22).

The heavy weight given to methods in the similarity relation makes it hard to conceive as a relation between worlds. Consider two worlds $w_1$ and $w_2$ and two subjects $S_1$ and $S_2$. Suppose that:

1. In $w_1$, $S_1$ believes $p$ on the basis of $m_1$ and $S_2$ believes $q$ on the basis of $m_2$. Both beliefs are true.

2. In $w_2$, $S_1$ believes $p$ on the basis of $m_1$ as well. $S_2$ believes $q$ but on the basis of $m_3$. Both beliefs are false.

Should $w_2$ count as close to $w_1$? If yes, we underestimate the position of $S_2$ at $w_1$; if no, we overestimate the position of $S_1$ at $w_1$. One solution would be to say that closeness is a contextual matter. If we are discussing the case of $S_1$, $w_2$ is close to $w_1$. If we are discussing the case of $S_2$, $w_2$ is more distant to $w_1$. But if we take this option, strengths of epistemic positions become context-sensitive: in a context in which we discuss $S_1$, the “strength” of $S_2$’s position is weaker (since the error world $w_2$ is “closer”) than if we discussed $S_2$. But DeRose explicitly requires a context-independent notion of strength of epistemic position (DeRose, 1995, 32–3).

There are two options here, I think. One is simply to relativise the error to be avoided to methods. We do not say that error about $p$ should be avoided at close worlds, where “close” implies that a similar method is used. Rather, we say that error about $p$ on the basis of the same method should be avoided at close worlds. Thus we can say that $w_2$ is a close world without underestimating the strength of $S_2$’s position, since $S_2$

---

19. The case is a variant of Nozick’s (1981, 179) grandmother case.
20. See Williamson (2000, 156–61) for further discussion.
avoids errors of the relevant type at $S_2$. The second option is to adopt a relation of closeness between cases instead of a relation of closeness between worlds. We have four cases: $(S_1, w_1), (S_1, w_2), (S_2, w_1), (S_2, w_2)$, the first is close to the second but the third is not close to the fourth.

I will adopt the first option here in order to keep DeRose’s semantics close to Lewis’s. As far as I can see the choice does not affect the present discussion.

6.3.3 The Rule of Sensitivity

DeRose gives one rule to indicate how the context selects a threshold for “knowledge”, the Rule of Sensitivity (1995, 36):

**Rule of Sensitivity** When a subject’s “knowledge” in $\alpha$ is under discussion in $C$, the threshold of $C$ is normally at least as high as to imply that $\alpha$ is case of “knowledge” only if it is a case of sensitive belief. (where $\alpha$ is a case of sensitive belief iff $\alpha$ is a case of belief, and at the closest worlds $w_\beta$ to $w_\alpha$ in which $p_\alpha$ is false, $S_\alpha$ does not believe $p_\alpha$.)

Suppose we discuss in $C$ whether one knows in $\alpha$. If the closest $p_\alpha$-world is distant, then satisfying “know” in $C$ requires a match of belief and fact up to distant worlds, i.e., a strong epistemic position. If the closest $p_\alpha$-world is close, then a weaker epistemic position may be sufficient to satisfy “know” in $C$.

The mechanism explains how sceptical arguments lead to true “knowledge” denials without showing that ordinary “knowledge” claims are false. A man asserts that he “knows” that he has hands. The proposition is false a close worlds; the epistemic position required for “knows” in his context is weak. A sceptic points out that if he “knows” that, then he “knows” that he is not having a perfectly realistic dream. The latter

---

21. It is not entirely clear how to cash out avoidance of ignorance on that option. What is “ignorance on the basis of a method”? We assume: one is ignorant on the basis of $m$ at $\alpha$ iff $p_\alpha$ is true, one uses $m$ at $\alpha$, but one does not believe $p_\alpha$ on the basis of $m$ at $\alpha$. (The “avoidance of ignorance” or adherence aspect of DeRose view will not bear on our discussion.)
proposition is false only at distant worlds. We thus get into a new context in which a stronger epistemic position is required for “know”. In the new context, the man truthfully asserts that he does not “know” that he has hands. A similar explanation can be given for the appeal of sceptical arguments in solitary thought. DeRose (1995, 36n) considers that the value of “know” could be affected otherwise than by discussing “knowledge”: for instance, attending to a proposition may suffice. However, it is not entirely clear how this would work, as we will see (6.6.4).

6.3.4 DeRose vs. Lewis on context shifts

Lewis’s and DeRose’s positions are quite close. On DeRose’s view, uttering that you “know” that you are not having a realistic dream normally forces the value of “know” to require you to avoid error in the closest world in which you are having a realistic dream. Since in that world, you are mistaken, the utterance is normally false. On Lewis’s view, the utterance makes you attend to the possibility that you are having a realistic dream, which in turn forces the value of “know” to require avoidance of error in (at least the closest version of) that possibility. So the utterance is false.

On Lewis’s view, context-shifts are pretty much automatic: when possible cases are brought to attention, they are thereby made relevant (see DeRose, 2009, 137–8). On DeRose’s view, context-shifts can be resisted (DeRose, 2009, 136–148). In solitary thought, if one reasserts to oneself that one “knows”, even though one realises one’s belief is insensitive, one may succeed in keeping the semantic value of “knows” low. What happens in cases of disagreement is more complex (DeRose, 2004; DeRose, 2009, 129–52). DeRose rejects the idea that disagreement divides the conversation into two contexts so that speakers in effect talk past each other (2009, 134–5). He also rejects both the view that the stricter value prevails (the sceptic wins) and the view that the looser prevails (the Moorean wins).

---

22. As DeRose (2009, 140) points out, this contrasts with Lewis’s (1979/1983, 245) view on contextual standards of precision, for which he does not endorse automatic context-shifts.
(2009, 136–41). He also rejects the view that the conversation determines a unique most reasonable standard that prevails (2009, 136–41), and the view that all “knowledge” claims become truthvalueless (2009, 143–4). His preferred view is that in such contexts “knowledge” claims are true iff they are true by the standards of all participants.  

6.3.5 The Rule of Sensitivity is specific to know, and has no clear analogue elsewhere in language

DeRose’s Rule of Sensitivity is specific to “know”, as Lewis’s rules are. It picks up, as a function of context and features of the modal space, a threshold on the scale of epistemic positions. The threshold is only relevant for “knows” and possibly other epistemic terms.

As far as I can see, the Rule of Sensitivity has no clear analogue elsewhere in language. Lewis (1996, 553) can draw on the analogy with quantifier domain restriction: which domain we quantify over with an utterance of (6.7), for instance, will depend on a number of features of the conversation and of the situation we are talking about.

(6.7) The French decided to leave.

Only very rough rules can be given: the salient people should be included (so that a salient Frenchman is not ignored); any person who is like the salient people in a salient respect should be included, and so on. Lewis’s idea is that analogous mechanisms of relevance govern “know”. But it is not easy to find analogues of DeRose’s Rule of Sensitivity.

First analogy: contrasting one’s case with clear counter-instances

Consider a first analogy that does not work. Suppose I am not sure whether I should consider myself “rich”. If my attention is drawn to the

23. The view is akin to supervaluationism about vague terms (Fine, 1975). The suggestion is that in conversations in which one participant is tempted by scepticism, the borderline area of “knows” expands so as to include all ordinary “knowledge” claims. On some of the precisifications acceptable in the context (those the sceptic has in mind), nothing is “known”, though by others, many things are “known”.
case of somebody who is much poorer than me, I will not be less tempted to consider myself “rich”. On the contrary, if anything, I will be more tempted to do so. Now suppose I consider myself to “know” that I have hands. My attention is then drawn to the possible predicament of René, who is deceived by an evil genius into thinking that he has hands. René’s epistemic position is much worse than mine: he has a false belief. So by analogous reasoning, I should, if anything, be more inclined to consider myself as “knowing” that I have hands.

Second analogy: contrasting one’s case with clear instances

A second unsuccessful analogy. Suppose again I am not sure whether I am “rich”. This time my attention is drawn to a person much richer than I am. I am then less inclined to consider myself as “rich”, if anything. Now suppose I consider myself to “know” that I have hands. My attention is drawn to whether I “know” that I am not dreaming. I am now considering a much stronger epistemic position, the thought goes, the position I would be in if I “knew” that I was not dreaming. So I am now less inclined to regard my position with respect to the fact that I “know” that I have hands. The line of thought is wrong: for all that has been said, my epistemic position with respect to the proposition that I am not dreaming may be just as good as my position with respect to the proposition that I have hands. Moreover, if it is better, the analogy predicts that I will be tempted to say that I may not “know” that I have hands, but that at least I “know” that I am not dreaming. Again, this is not what DeRose wants.

Third analogy: threshold suggestion

Another analogy would be that of threshold suggestion. Suppose I am not sure whether I am “tall”. A certain door is brought to my attention. This may tempt me to use the door as a point of reference for “tallness”. Here I do not compare myself to instances that are higher or lower on the scale than I am. Instead, a possible threshold is made salient to me. Similarly, when the proposition that \( p \) is brought to attention, there may
be a temptation to use the distance at which \( p \) has a different truth-value as the threshold.

The analogy is quite unsatisfactory, however. Threshold suggestion, if it exists at all, is a weak effect. Consider:

(6.8) Many tall Danes were looking at the Eiffel tower.

The fact that the Eiffel tower is mentioned does not make it harder to say that the Danes in question are “tall”.\(^{24}\) Moreover, threshold suggestion does not appear to be upwards-directed. That is, looser thresholds are as easily suggested as stricter ones. But in the “knowledge” case, lowering thresholds in a conversation is hard.

**Fourth analogy: assimilating one’s case to counter-instances**

Consider now a case that better mirrors the “knowledge” one. Suppose I consider myself “healthy”. My attention is brought to somebody who appeared just as “healthy” as me but had a sudden heart attack. I am then more hesitant to consider myself “healthy”. Similarly, when my attention is brought to some subject who takes himself to be in just as good an epistemic position as I am, but who is not, I am more hesitant to consider myself as “knowing”. In both cases, pointing out that my case resembles a clear case of non-“healthiness” or non-“knowledge” makes me reluctant to apply the predicate to my own case.

The analogy is good, but it will not help DeRose. For in such a case, I am not raising the threshold for “healthiness”, I am merely starting to doubt whether I satisfy it. I can be led to say “Well, perhaps I am not healthy, then”. But I will not be lead to say, “Well, I am not healthy, then”. In the “knowledge” case, I would not be requiring a stronger epistemic position for “know”. I would only be starting to doubt whether the sceptical possibility is closer than I thought. I could reply to the sceptical

\(^{24}\) For the purposes of the analogy, it cannot be objected that threshold suggestion does not work in that case because nobody could be as tall as the Eiffel tower. For DeRose contends that sceptical arguments can set thresholds that nobody can meet.
argument by “Well, perhaps I do not know, then.” But not by “I do not know, then.” 25

The latter analogy in fact suggests that if speakers follow the Rule of Sensitivity, they are confusing signs of “knowledge” and “knowledge”. Suppose I consider myself “rich” and I am independently certain that my neighbour is as rich as I am. I have a fancy car, and I take this to be a mark of being “rich”. Upon realising that my neighbour does not have a fancy car, I conclude that he is not “rich”, and infer that I am not “rich” either. The reasoning seems blatantly distorted, but it is strictly analogous to what DeRose thinks goes on in the case of “know”. (Sensitivity is represented by having a fancy car, and the neighbour represents the belief that some sceptical hypothesis does not obtain.) I consider myself to “know” that I have hand. The belief is sensitive, and I take that to be a mark of “knowledge”. I am also certain that my belief that I have hands is not better “knowledge” than my belief that I am not dreaming. Upon realising that the latter belief is not sensitive, I conclude that it is not “knowledge”, and infer that I do not “know” that I have hands. In doing so, I would be confusing the real measure of knowledge — the avoidance of error over a sufficiently wide area — with a mere sign of it — sensitivity. 26

On that view the conclusion that I do not “know” is simply false. This is not the view DeRose wants.

To sum up the issue, DeRose takes it as manifest that there is a tendency to deny “knowledge” when insensitive belief are presented. He thinks it implausible that we are wrong in such cases, so rejects the kind of diagnosis Sosa proposes. He also thinks it implausible that we are wrong in our judgements of comparative epistemic strength, which would be the case if sensitivity was required for “knowledge”. 27 So he concludes that we

25. See Sosa (2000b, 4) for a related idea.
26. On Sosa’s view the confusion between the two is not due to a confusion between a sign of knowledge (sensitivity) and its real measure (safety), but to a confusion between the sensitivity conditional and its converse.
27. An example of the judgements in question is: “if he knows that he has hands, then
shift the value of “know” in context so that it appears as if “knowledge” required sensitivity. The corresponding contextual Rule of Sensitivity seems to have no analogue elsewhere in language.

6.3.6 Formal version of DeRose’s view

Let me finally state DeRose’s “rough and ready” semantics formally. Write:

\( Bm\alpha \) In \( \alpha \), \( S_\alpha \) believes that \( p_\alpha \) on the basis of \( m \).

\( d(w, w') \) is the distance between \( w \) and \( w' \). For any \( w, w', d(w, w') \in [0, 1] \). 0 is total similarity (\( d(w, w) = 0 \) for all \( w \)), 1 is maximum dissimilarity.

\( E_{m\alpha} \) In \( \alpha \), \( S_\alpha \) has a false belief on the basis of \( m \): \( Bm\alpha \land \neg T\alpha \).

\( I_{m\alpha} \) In \( \alpha \), \( S_\alpha \) is ignorant that \( p_\alpha \) on the basis of \( m \). (This can be made explicit has follows: in \( \alpha \), \( p_\alpha \) is true, \( S_\alpha \) uses \( m \), but \( S_\alpha \) does not believe \( p_\alpha \).)

\( k_C \) the threshold of epistemic strength selected by context \( C \). For any \( C \), \( k_C \in [0, 1] \).

\( R_C \) the accessibility relation between worlds resulting from the threshold \( k_C \): for any worlds \( w, w' \) and context \( C \), \( wR_Cw' \) iff \( d(w, w') \leq k_C \).

**DeRose’s (rough) Modal Contextualism** For all \( C, \alpha \), \( K_C\alpha \) iff for some \( m \):

\( Bm\alpha \), and

\( \neg E_{m\beta} \) and \( \neg I_{m\beta} \) for all \( \beta \) such that \( w_\alpha R_Cw_\beta \) and \( p_\beta = p_\alpha \).

For all \( C, \alpha \), \( K_C\alpha \) iff \( \exists m(Bm\alpha \land \forall \beta((w_\alpha R_Cw_\beta \land p_\beta = p_\alpha) \rightarrow (\neg E_{m\beta} \land \neg I_{m\beta}))) \).

**Rule of Sensitivity** If \( \alpha \) is under discussion in \( C \), then \( k_c \) is normally such that \( k_C \geq d(w_\alpha, w_\beta) \) where \( w_\beta \) is one of the closest worlds to \( w_\alpha \) were \( p_\alpha \) is false: \( \forall w_\gamma(w_\gamma \models \neg p_\alpha \rightarrow d(w_\alpha, w_\gamma) \geq d(w_\alpha, w_\beta)) \).

---

he knows that he is not dreaming."
6.4 The necessary truths problems (again)

Lewis’s and DeRose’s contextualist modal requirements are proposition-centred. To satisfy “know” \( p \) in a context \( C \), one should avoid error \( \text{that } p \), or error \( \text{that } p \text{ on a given basis} \), over a certain range of possible cases. As a result, the requirements are trivially satisfied by beliefs in necessary truths. We have encountered the problem in sec. 4.1.1.

DeRose does not discuss the problem, as far as I am aware. He might assume that it is dealt with by some other condition on knowledge than the modal strength of one’s epistemic position. Or he might assume, following Nozick, that it is solved by his requirement of avoidance of ignorance. We have seen that both solutions are flawed. Here we examine alternatives put forward by Lewis’s and those using Lewis’s framework.

6.4.1 Lewis’s “linguistic disguise” suggestion

Lewis acknowledges the problem, but sets it aside as an issue of “linguistic disguise” of propositions:

So the necessary proposition is known always and everywhere. Yet this known proposition may go unrecognised when presented in impenetrable linguistic disguise, say, as the proposition that every even number is the sum of two primes. Likewise, the known proposition that I have two hands may go unrecognised when presented as the proposition that the number of my hands is the least number \( n \) such that every number is the sum of \( n \) primes. […] These problems of disguise shall not concern us here. Our topic is modal, not hyperintensional, epistemology.

We can see how “disguises” of propositions allow one to simultaneously believe and fail to believe a proposition. But it is hard to see how to apply

---

28. We argue in 4.1.1 that if another condition solves the problem, then it will likely show sensitivity to be redundant or an under-generalisation. That Nozick’s Adherence does not solve the problem is argued in sec. 4.1.2.
the idea to knowledge. Take our Prime Numbers case (p. 156): Primo believes rightly that 47 is prime, but on the basis of a mistaken calculation that would lead him to mistakenly believe that 49 is prime. Assume Primo knows that $2 + 2 = 4$. On Lewis’s coarse view of propositions, the proposition that $2 + 2 = 4$ and the proposition that 47 is prime are the same, so Primo knows that 47 is prime. Yet, he does not know that 47 is prime under that disguis "47 is prime”. Why not? He is certainly aware of the proposition under that disguise, that is, he can think it under that disguise.  

Lewis suggests that what Primo lacks is having recognised that the proposition 47 is prime, under that disguise, is the same as the proposition that $2 + 2 = 4$, under that disguise — or more generally, the same as some proposition, under a disguise, such that he knows that proposition under that disguise. If recognising is knowing, what Primo lacks is to know that $2 + 2 = 4$ is equivalent to 47 is prime, under that disguise. The proposition that $2 + 2 = 4$ is equivalent to 47 is prime is again the necessary proposition. So we are back with a version of the same question: why does Primo not know that $2 + 2 = 4$ is equivalent to 47 is prime, under that disguise? If we go on saying that he lacks the knowledge that $2 + 2 = 4$ is equivalent to the proposition that $2 + 2 = 4$ is equivalent to 47 is prime, we get into a regress. If recognising is not knowing, I am not sure what it is.

The problem gets worse if we focus on pairs of propositions which are not equivalent, as in the rigid / non-rigid variants of the fake barn case (p. 159). On Lewis’s coarse view of propositions, Bernard knows that that building is a barn, but not under that disguise. He does not know, however, that the building he is looking at is a barn, under any disguise. Lewis has to claim that the two failures of knowledge belong to different kinds: one is a failure in recognising the necessary truth under a specific linguistic disguise, the other is a failure in avoiding possible errors. The

---

29. Awareness models of non-logically omniscient agents (Fagin and Halpern, 1988) face the same problem: as long as a formula for a necessary truth is in the set of formulas the agent is “aware of”, the formula is known.

30. I am not sure whether the regress is vicious in any way, but it is certainly not explanatory.
first is in the province of “hyperintensional” epistemology, the second in the province of “modal” epistemology. But it is incredible that such a difference should turn on Bernard’s choice of a referring device.\footnote{Let me note two related problems for Lewis’s account. (1) The requirement is trivially satisfied by propositions concerning one’s experience (\textit{Hawthorne}, 2004): if I have a visual experience as of 157 phenomenal dots, I know that I have an experience as of 157 phenomenal dots. (Lewis could answer that I do, but not under that disguise.) (2) Lewis’s account validates a \textit{S}5 epistemic logic (Lloyd Humberstone quoted by \textit{Williamson}, 2001, 29), which implies that whenever I do not know that \( p \), I know that I do not know that \( p \). (If it is possible that I have experience \( e \) while \( p \) is false, then it is necessary that it is possible that I have experience \( e \) while \( p \) is false, so in all possibilities in which I have experience \( e \), my experience is \( e \) and it is true that it is possible that I have experience \( e \) while \( p \) is false. By Lewis’s definition, it follows that when I have experience \( e \), I know that I do not know \( p \).) Again Lewis could again answer: I do, but not under that disguise.}

### 6.4.2 The proper basing answer

Followers of Lewis (Blome-Tillmann, 2009b, 259–60, Ichikawa, ms.b) have tried to avoid the problem by introducing an additional condition on knowledge.\footnote{This contravenes the Simplicity rule we have defended sec. 4.4.4.} Blome-Tillmann’s (2009b, 259–60) proposal is that the belief should be “properly based”:

**Proper basing** For all \( \alpha, C \), \( K_C \alpha \) only if in \( \alpha \), \( S_\alpha \) believes that \( p_\alpha \) and \( S_\alpha \)’s belief that \( p_\alpha \) satisfies “is properly based” in \( C \).\footnote{We have to state the proper basing condition in metalinguistic terms because, as we will see, Blome-Tillmann’s formulations have readings which entail that whether a belief is “properly based” varies with context.}

Blome-Tillmann adds that a belief based on an unreliable method is not properly based (260). But, as he notes, having a generally reliable method cannot be 	extit{sufficient} for proper basing, on pain of Gettier counterexamples. Suppose \( S \) has a false belief that \( p \) based on a reliable method and deduces that \( p \lor q \), where \( q \) is a necessary truth; \( S \)’s belief that \( p \lor q \) would be necessarily true and properly based, and thus wrongly counted as knowledge (260n). Blome-Tillmann’s solution is to provide an explanationist gloss of proper basing:

**Proper basing (2)** For all \( \alpha, C \), if \( \alpha \) is a belief case, \( S_\alpha \)’s belief that \( p_\alpha \) satisfies “is properly based” in \( C \) iff \( S_\alpha \) holds the belief that \( p_\alpha \) for
the reason that her evidence eliminates the not-\(p_\alpha\) possibilities that cannot be properly ignored in \(C\). ({Blome-Tillmann, 2009b}, 260n)

The gloss turns “proper basing” into a context-sensitive matter: a belief may be rightly called “properly based” in a context \(C\) while being properly called “improperly based” in another, since “proper basing” depends on the set of possibilities relevant in a context. “Reason” could be understood as a mental state of \(S_\alpha\) or as an objective explanation ([Blome-Tillmann, 2009b], 260n). But on any of these readings, “the not-\(p_\alpha\) possibilities that cannot be properly ignored in \(C\)” cannot take narrow scope under “reason”. In the current context, the following is true: “[Just before he died] Caesar knew that Brutus had stabbed him”. Yet the fact that some possibilities are ignored in our current context was neither part of Caesar’s own reasons to believe that Brutus stabbed him, nor part of the explanation of why he formed that belief. So “the not-\(p_\alpha\) possibilities […]” should take wide scope or directly refer to the possibilities in question. To disambiguate:

**Proper basing (3)** For any \(C\), let \(P_C\) be the set of possibilities that cannot be ignored in \(C\).

For all \(\alpha, C\), if \(\alpha\) is a belief case, \(S_\alpha\)'s belief that \(p_\alpha\) satisfies “is properly based” in \(C\) iff \(S_\alpha\) holds the belief that \(p_\alpha\) for the reason that her evidence eliminates \(\neg p_\alpha\)-possibilities in \(P_C\).

On the mental state reading, this requires Caesar to think of \(\neg p\)-possibilities in \(P_C\). The reading is extremely implausible: it is dubious that Caesar has or needs to have thoughts about the possibilities open in our context or his own evidence, explicitly or implicitly.\(^{34}\) The objective explanation

\(^{34}\) First it is utterly implausible that Caesar could think that his evidence eliminates all the possibilities that are in the set \(P_C\) of the current context. One of them is that Barack Obama is reelected, for instance. Moreover, even if we somehow restricted ourselves to possibilities relevant to his assassination, Caesar need not have thought of the very same possibilities. For instance, where we think of the possibilities that the conspirators hid daggers under her toga, he may have considered the possibility that Casca carries a dagger and the possibility that Brutus carries a dagger and so on. One may try to weaken the requirement by saying that Caesar should have thought of possibilities whose union coincided with the set of possible worlds corresponding to the open possibilities of \(C\), but I see little hope that way either.
6.4. The necessary truths problems (again) 303

reading is tantamount to, simply:

For all \( \alpha, C \), if \( \alpha \) is a belief case, \( S_{\alpha} \)'s belief that \( p_{\alpha} \) satisfies “is properly based” in \( C \) iff \( S_{\alpha} \) believes that \( p_{\alpha} \) because her evidence eliminates \( \neg p_{\alpha} \)-possibilities in \( P_{C} \).

There are two major problems with the suggestion. First, it is not clear what its application to necessary truths is. What is it for someone to believe that \( 2 + 2 = 4 \) because her evidence eliminates cases in which \( 2 + 2 = 5 \), for instance? The problem is that the phrase “\( \neg p \)-possibilities in \( P \)” presupposes the existence of such possibilities, while if \( p \) is a necessary truth, there are none. We are essentially thrown back to a variant of the necessary truth problem in which the requirement is inapplicable rather than trivially satisfied, as with Nozick’s sensitivity. Second, if there are many different sets of possibilities \( P_{C'} \) for various contexts \( C' \), the proper basing clause will too often fail as a result of the lack of generality of the explanandum. Suppose that I believe that a tomato is red because it is red. Then it is false that I believe it because it is of this or that particular shade of red, or because it is of a particular subset of the shades of red, and so on. Similarly, if \( S_{\alpha} \) believes that \( p_{\alpha} \) because her evidence eliminates \( p_{\alpha} \)-possibilities in \( P_{C} \), it is likely to be false that she believes that \( p_{\alpha} \) because her evidence eliminates \( p_{\alpha} \)-cases in \( P_{C'} \), where \( P_{C'} \) is a subset of \( P_{C} \). So ignoring more possibilities in \( C' \) may make \( C' \) more stringent than \( C \), since it would not be true in \( C' \) that \( S_{\alpha} \)'s belief is “properly based”. This is implausible. 35, 36

35. More generally, Blome-Tillmann seems to assume than necessary truths are a special epistemic case: “my view still has as a consequence that having one’s belief properly based is sufficient for ‘knowing’ necessary truths. Even though this consequence of PEC may initially seem implausible, I do not think that it is mistaken: regarding necessary truths there simply are no epistemically deficient ways to believe apart from those involving elements of improper causal sustenance” (Blome-Tillmann, 2009b, 260). But as we argued sec. 4.1.1, the assumption is unwarranted.

36. Another type of reply is to require that the subject bases her belief on her evidence. On Lewis’s internalist notion of evidence as experience, the condition is too weak. In the rigid fake barn case (p. 159), the subject bases her belief in a necessary proposition on her evidence. Ichikawa (ms.a) suggests instead giving up Lewis’s internalist notion of evidence and endorsing Williamson’s (Williamson, 2000, ch.9) claim that evidence is knowledge. Lewis’s semantics becomes circular, but that is acceptable. The proposal
6.4.3 Methods contextualism

The solution is, as I have argued, to replace proposition- or belief-centred requirements with method-centred ones. Lewis’s notion of evidence is too coarse for that purpose: we do not want to deny knowledge as long as one could have some false belief while having the experience one has. But let us assume an adequate notion of method can be used (see sec. 4.1.2). Let us also leave aside DeRose’s avoidance of ignorance requirement, and let us require belief for knowledge (contra Lewis). Write:

\[ E_{m\alpha} \text{ in } \alpha, S_\alpha \text{ believes } p_\alpha \text{ on the basis of } m \text{ but her belief is false: } B_{m\alpha} \land \neg T_\alpha. \]

We propose:

**Generic methods contextualism, Lewis-style** For all \( C, \alpha, K_C \alpha \iff \text{for some method } m: B_{m\alpha} \text{ and for all } \beta \text{ such that } \alpha \in P_C, \neg E_{m\beta}, \)

where \( P_C \) is the union of possibilities that are not properly ignored in \( C \).

**Generic methods contextualism, DeRose-style** For all \( C, \alpha, K_C \alpha \iff \text{for some method } m: B_{m\alpha} \text{ and for all } \beta \text{ such that } \alpha R_C \beta, \neg E_{m\beta}. \)

where \( \alpha R_C \beta \iff \text{the distance between } w_\alpha \text{ and } w_\beta \text{ satisfies the threshold of } C: d(w_\alpha, w_\beta) \leq k_C. \)

On the method amendment, the strength of one’s epistemic position in a belief case \( \alpha \) is given by the distance at which one finds the closest world in which the best method on which the belief is based yields a false belief.\(^{37}\)

The methods amendment gives a partial reply to a problem for DeRose’s contextualism raised by Blome-Tillmann (2009a). Blome-Tillmann (2009a, 387) discusses the following dialogue:

\((6.9) \text{ Vladimir and Gogo:}\)

\[ G: \text{Is it true that nothing can travel faster than light?}\]

implies that a belief in a necessary truth that is based on “knowledge” is “knowledge”. This is still not sufficient to solve fake barn cases. (A similar problem affects no-false-assumptions approaches; see Lycan, 2006.)

\(^{37}\text{ We allow that a belief be based on many methods. The best one is one whose closest error is the furthest away.}\)
6.4. The necessary truths problems (again)

V: Yes, that’s true.
G: I didn’t know that. (silence) but Vladimir...
V: What is it?
G: I know that that’s a zebra! See the black and white stripes?
V: (angrily) Everybody knows that that’s a zebra!

Blome-Tillmann (388) argues that DeRose’s view cannot yield the desired verdict for some such cases, namely that all the utterances come out true. By the Rule of Sensitivity, the third utterance of Gogo raises the threshold for “know” to avoidance of error at all worlds up to the closest worlds in which something travels faster than light. Many bizarre worlds are closer than this, including ones in which Gogo is mistaken about the zebra. So the last utterance turns out false.

The partial reply is the following. On DeRose’s view, the Rule of Sensitivity is not automatic. So it is open for Vladimir and Gogo not to raise requirements for “know” so that sensitivity with respect to the proposition that nothing can travel faster than light is required. Still, Gogo’s third utterance may be true, either because he fails to believe the proposition, or because he believes it on the basis of a method that would yield false beliefs in other propositions. No physically impossible worlds need to be brought into consideration.

The reply is partial, because it does not explain why the Rule of Sensitivity does not operate in that context. A natural idea here would be to replace the Rule of Sensitivity with a method-based analogue, in the hope that the latter does not apply to that case. But it is hard to see how that would go. Sensitivity makes essentially reference to a particular proposition, and considers what would have happened if that proposition had been false. A method-based conditional such as “if one was to believe something false on the basis of a method, it would not be on the basis of m” does not seem to make sense here. I will not explore the option further.

Lewis’s contextualism faces a similar problem. Presumably, speakers in (6.9) are attending to the possibility that something travels faster than light. It is doubtful that all sub-cases, or even all close sub-cases, of this
possibility will be eliminated by Vladimir’s experience. So there is a risk that it turns out false that Vladimir “knows” that nothing travels faster than light in any context in which that is asserted.

6.5 Attending, presupposing and taking seriously

6.5.1 Attending is not sufficient to take seriously

As a piece of empirical linguistics, Lewis’s Rule of Attention is false: thinking of a sceptical scenario does not automatically sway people’s intuitions in a sceptical direction, as Hawthorne (2004, 64) notes:

I go to the movies and see The Matrix, a modern brain in a vat tale about how humans are really organisms sustained in pods that generate hallucinations. I can watch the movie without taking seriously the idea that I am a pod person (even if I notice that it is true in the story that I am a pod person). 38

One could still maintain that the possibility becomes relevant, even though it is not taken seriously, and that subsequent utterances of “I know that I am in a movie theatre” are false, contrary to the speaker’s intuitions. But as Hawthorne notes (64), that is implausible. It would undermine the motivation for contextualism.

DeRose does not accept anything as strong as the Rule of Attention. On his view, the Rule of Sensitivity represents only a tendency of speakers. Most agree that sceptical intuitions kick in when sceptical possibilities are taken seriously, but it is unclear what taking seriously amounts to. 39

38. Hawthorne credits Alyssa Ney with the remark.
39. See Hawthorne (2004, 168–70). Hawthorne’s discussion is on two (tentative) assumptions: (a) taking sceptical considerations seriously does not affect the value of “know” in the conversation of those who do take them seriously, but it destroys their own knowledge, and (b) taking a sceptical hypothesis sh seriously consists in the “intellectual seeming” to one that sh is an epistemic possibility for one; in other terms, it seems to one that one does not know ¬sh. Hawthorne then asks how the fact that it seems to one that one does not know can bring about that one does not know ¬sh. He
6.5.2 The no-belief account of taking seriously

Blome-Tillmann (2009b, 248) tries to say a bit more, by replacing Lewis’s Rule of Attention with a Rule of Presupposition:

Rule of Presupposition For any C, α, if \( w_n \) is compatible with the pragmatic presuppositions in C, α cannot be properly ignored in C.  

(Note that the Rule only replaces the Rule of Attention; Lewis’s other Rules are still in place, and may force into relevance possibilities that speakers presuppose not to obtain.) Drawing on Stalnaker (1999, 2002), Blome-Tillmann characterises pragmatic presupposition as follows (256):

Pragmatic Presupposition S pragmatically presupposes \( p \) in C iff S is disposed to behave, in her use of language, as if she believed \( p \) to be common ground in C, where \( p \) is common ground in C iff all members of C accept (for the purpose of conversation) that \( p \), and all believe that all accept that \( p \), and all believe that all believe that all accept it, and so on (250).

A possibility (proposition) \( p \) is taken seriously in a context iff nothing is presupposed in that context that entails that \( p \) is false.  

Blome-Tillmann (2009b, 248) extends pragmatic presupposition to the realm of thought. The common ground in that case is the set of prop-

40. The rule should be generalised to cases instead of worlds, since we can presuppose de se propositions. For instance, if I suffer from amnesia, I may presuppose that I am French without presupposing that I am Julien Dutant. Cases in which the subject is not French can be ignored, but cases in which some other French person has the experiences I have should not.

41. Some details have to be refined. Georges pragmatically presupposes that God exists; however, he takes seriously the possibility that there is no matter. Suppose that, as a matter of metaphysical necessity, there is no God. It follows that no possible world \( w \) is compatible with Georges’s presuppositions. Thus his taking seriously the possibility that there is no matter would have no effect whatsoever on the context. But that is implausible.
positions that $S$ accepts (for the purpose of inner “conversation”), believes herself to accept, believes herself to believe to accept, and so on — call that set the personal ground of $S$. There are two important questions about the personal ground that Blome-Tillmann does not raise explicitly: (1) Can one accept some proposition $p$ for the purpose of mental speech without believing $p$? and (2) Can someone behave in mental speech as if $p$ was part of the personal ground without $p$ being part of the personal ground? If both answers are “no”, the personal ground collapses into the agent’s beliefs.

It is important to note here that, as Blome-Tillmann understands it, pragmatic presupposition is distinct from mere assumption (2009b, 278). If a speaker utters “We all know it’s false, but let’s assume that pigs can fly”, and the audience complies, it is assumed in the subsequent conversation that pigs can fly. However, it is not presupposed, as evidenced by the participants’s retaining their dispositions to answer “Of course they can’t — but we assume for the moment that they can”, if asked whether pigs can really fly, or to utter claims such as “I’m glad that pigs can’t fly”. 42

With the distinction in place, I cannot imagine somebody pragmatically presupposing something in solitary thought without believing it. One may utter to oneself, in thought, “The king of France is bald!”, or imagine oneself speaking to a XXIth century king of France and saying to him “Your Majesty is bald”. But these are not genuine mental assertions and they do not genuinely presuppose the existence of a king of France. One may also assume the existence of a king of France in hypothetical reasoning, but as Blome-Tillmann points out, assumption is not presupposition. Perhaps in some cases of self-deception one does assert to oneself things that one does not believe. A hard-nosed monarchist may manage to pre-

---

42. One may contrast assumption and presupposition as follows. Assumption is a disposition to accept a proposition for some subpart of one’s current linguistic behaviour — together with a belief that all accept it for that purpose, a belief that all believe that all do, and so on — while presupposition proper is acceptance for all of one’s current linguistic behaviour — and a belief that all do, and so on. This requires, however, that there is a sense in which the presupposed propositions are accepted even in those sub-parts of the discourse in which their acceptance is apparently suspended, as is the case of the presupposition that pigs do not fly in Blome-Tillmann’s example.
suppose in thought the present existence of a French king without really believing in it. But barring these special cases, it does not seem possible to pragmatically presuppose something that one does not believe.

The upshot is that if we focus on the simple cases of solitary thought and soliloquy, Blome-Tillman’s suggestion is that taking \( p \) seriously is simply \textit{not believing that not-}\( p \). (More precisely, not believing something that is incompatible with \( p \).) The view is attractive. If I now believe that I am alone in my flat, then I am not currently taking seriously the possibility that there is someone else in the flat. Of course if I were to hear a noise, I may come to take it seriously; but right now, I am not. If I do \textit{not} believe that I am alone (whether or not I would believe that I am \textit{not} alone), then I take the possibility seriously. To take it seriously, it is not necessary that I am currently thinking of it or worrying about it. It is sufficient that I am at least currently disposed to take it into account in practical plans.

### 6.5.3 The problem of third-person sceptical arguments

The no-belief account of taking seriously runs into a problem with \textit{third-person sceptical arguments}, however. Consider the following two dialogues derived from Dretske’s zebra case. In the first, Bouvard and Pécuchet are at the zoo looking at a zebra:

\begin{align*}
4.10) & \quad \text{Bouvard. I know that that’s a zebra.}
\end{align*}

\footnotesize
\begin{itemize}
\item 43. As I imagine the case, the monarchist believes, or rather tries to believe, of some actual person that, in virtue of some inheritance rules, that person is the present legitimate king of France. By uttering “the king of France” in thought and speech, she intends to refer to that person. But at the same time, she cannot help but be aware of the fact that that person is not really a king (i.e., he is not a chief of state nor close to being one). Perhaps that makes her accept the proposition that that person is king of France in thought without believing it. I let the reader decide.
\item 44. At least when formulated with a notion of \textit{outright belief} as opposed to qualified belief. Outright belief is a categorical form of belief, qualified belief is typically represented as probabilistic belief. Believing outright does not require believing with certainty or firmness. It is rather believing without question or doubt. Things that we take for granted without even thinking of them (e.g. that a certain ice cube will be cold to touch) are perhaps the best illustration of what is intended by the notion.
\item 45. One easily imagines the main characters of Flaubert’s \textit{Bouvard et Pécuchet} holding such dialogues.
\end{itemize}
Pécuchet. Well, it might be a mule cleverly painted to look like a zebra! You would not be able to tell the difference.

B. You’re right! I don’t know whether it is a zebra after all.

In the second, Bouvard and Pécuchet fancy themselves as zebra connoisseurs and have no doubt about the species of the animal in the pen. They are looking at a visitor looking at the zebra:

(6.11) Bouvard. She knows that that’s a zebra.

Pécuchet. Well, it could have been a mule cleverly painted to look like a zebra! She would not be able to tell the difference.

B. You’re right! She doesn’t know whether it is a zebra after all.

Each argument is equally effective in leading one to accept its conclusion. (Which is not to say that they always succeed.) Of course the first argument is more unsettling to Bouvard. He ends up thinking that he does not know that the animal is a zebra. Presumably, he has also ceased to believe that it was a zebra, and is now wondering whether it is one or not. In the second case, by contrast, Bouvard need not reassess his convictions. He can rest assured that the animal is a zebra and that he knows it. Still, the possibility raised by Pécuchet convinces him that the visitor does not know. As far as ascribing knowledge goes, the arguments are equally efficient, though the conclusion of the first is more unsettling to one who reaches it.

Both arguments can be reinforced by a use of deductive closure. We first claim that the error possibility itself is not known not to obtain, and then claim that a proposition that entails that it does not obtain cannot be known either:

(6.12) Bouvard. I know that that’s a zebra.

Pécuchet. Well, it might be a mule cleverly painted to look like a zebra! Do you know whether it is not a cleverly painted mule?

B. No, I guess I don’t.

P. If you don’t know whether it is not a cleverly painted mule, how can you know that it is a zebra?
B. Well, I guess I can’t.

P. So you don’t know that it is a zebra, after all?

B. You’re right, I don’t. 46

(6.13) Bouvard. She knows that that’s a zebra.

Pécuchet. Well, it could have been a mule cleverly painted to look like a zebra! Does she know that it is not a cleverly painted mule?

B. No, I guess she doesn’t.

P. If she doesn’t know that it is not a cleverly painted mule, how can she know that it is a zebra?

B. Well, I guess she can’t.

P. So she doesn’t know that it is a zebra, after all?

B. You’re right, she doesn’t.

The third-person arguments are even more efficient if Pécuchet has some ground to claim that there could have been a painted mule instead of the zebra:

(6.14) Pécuchet. Well, it could have been a mule cleverly painted to look like a zebra! The zoo’s mules have been taken out this week to refresh their paint, but they normally put some in the pen along with the zebras.

The intuitive appeal of third-person sceptical arguments cannot be accounted for by Blome-Tillman’s presupposition view, nor by the idea that sceptical arguments work through removing belief. 47 In the third-person dialogues (6.11, 6.13), both speakers presuppose that the animal in the pen is not a painted mule. It is perfectly felicitous for either of them to say: “she doesn’t know that the zebra is a zebra”, “the zebra looks like a painted mule”, and so on. Correlatively, both believe that the animal is

---

46. In this dialogue, as opposed to the next one, it is more natural to use “know whether” formulations than “know that” ones, because the latter presuppose the truth of the embedded proposition, and the point of the present dialogue is precisely to remove the presupposition. I do not see that this affects the parallel between the cases.

47. The point holds whether one takes the presupposition-removal / belief-removal story as a contextualist account of “knows” or merely as a psychological account of the appeal of sceptical arguments.
a zebra, and it is common ground among them that they do so. So either a possibility of error can lead to a denial of knowledge even though it is not taken seriously, or a possibility of error can be taken seriously even though the speakers believe that it does not obtain.

Blome-Tillman may reply by saying that the possibility remains irrelevant in the dialogues (6.11, 6.13), and thus that Bouvard is mistaken in accepting the conclusion. That undermines his contextualist treatment of the arguments (6.10, 6.12), which is motivated by the idea that speakers’s intuitions are right in such cases. Alternatively, he can reply that the painted-mule possibility in cases (6.11, 6.13) is made relevant by another rule than the Rule of Presupposition — presumably, by the Rule of Resemblance. But this would undermine the need for the Rule of Presupposition to account for the other cases.

6.6 The unmentioned subjects problem

Lewisean contextualist accounts face a problem arising from unmentioned subjects. As far as I know, the problem has not been pointed out in the literature. Yet it has important consequences: it shows that Lewis-style contextualist semantics violate the factivity of knowledge — the idea that only truths can be “known”. DeRose-style semantics avoid the problem, but the considerations behind it raise a puzzle for the application of his Rule of Sensitivity.

6.6.1 Semantics for unuttered sentences

In standard semantics, semantic values are not only ascribed to uttered expressions, but also to unuttered ones. This is crucial to give a compositional semantic account of sentences involving operators or quantifiers. To illustrate, consider an utterance of (6.15) on a particular boat:

(6.15) Someone on the boat is happy.

Suppose Isaac is the only happy person on the boat. Then (6.15) is true in its context of utterance because Isaac satisfies “is happy” in the context
of utterance, even if Isaac is not mentioned in that context. Similarly, consider an utterance of (6.16):

(6.16) Isaac could have been happy.

If the utterance is true, it will be so in virtue of there being a relevant world \( w \) and time \( t \) at which “Isaac is happy” is true in the context, even if that world and time are not mentioned in the context.

### 6.6.2 Factivity fails in Lewis’s semantics

Now consider an actual context \( C \) in which it is true to utter:

(6.17) Alice knows that she has hands.

Let @ be the actual world. Leaving time aside, by Lewis semantics we have:

1. Alice satisfies “knows that she has hands” in \( C \) at @ iff in all worlds \( w \) in \( D_C \), if Alice’s evidence in \( w \) is the same as Alice’s evidence in @, Alice has hands in \( w \).

Where \( D_C \) is the set of worlds that are not properly ignored in \( C \). Now consider a world \( w^* \) in which Alice has the evidence she actually has (that is, the same experience and memories, in Lewis’s use), but where she is a handless victim of an Evil Demon. Crucially, the possibility that a world like \( w^* \) obtains is not mentioned in \( C \). We have:

2. Alice satisfies “knows that she has hands” in \( C \) at \( w^* \) iff in all worlds \( w \) in \( D_C \), if Alice’s evidence in \( w \) is the same as Alice’s evidence at \( w^* \), Alice has hands in \( w \).

Given that Alice’s evidence in \( w^* \) is the same as her evidence in @, we have:

3. Alice satisfies “knows that she has hands” in \( C \) at \( w^* \) iff in all worlds \( w \) in \( D_C \), if Alice’s evidence in \( w \) is the same as Alice’s evidence at @, Alice has hands in \( w \).

By (1) and (3) we have:

---

48. That is, the union of the set of possibilities that are not properly ignored in \( C \).
(4) Alice satisfies “knows that she has hands” in C at $w^*$ iff Alice satisfies “knows that she has hands” in C at @.

By assumption, Alice satisfies “knows that she has hands” in C at @. Moreover, Alice is handless at $w^*$, so “Alice does not have hands” is true at $w^*$ in C. So by (4) we have:

(5) “Alice does not have hands and Alice knows that she has hands” is true at $w^*$ in C.

Which contradicts the factivity of knowledge.

Now given the Rule of Actuality, if the case of Alice in $w^*$ came under discussion, we would get in a context $C^*$ in which $w^*$ could not be properly ignored, and Alice would not satisfy “knows that she has hands” in $C^*$ at $w^*$ — nor at @, for that matter. So it seems that failures of factivity at least cannot be asserted. Even if that was so, however, the problem would be bad enough. It is mysterious why we would use “know” in some context to express a property that brain-in-vats have with respect to false propositions.

The problem gets worse, however, under two assumptions:

(6) “could have $p$” is true in C at $w$ iff there is a world $w'$ such that “$p$” is true in C at $w'$ and $w'$ satisfies “could have obtained” in C at $w$.

(7) $w^*$ satisfies “could have obtained” in C at $w$.

By (5), “Alice knows something false” is true in C at $w^*$. By (7) and (6), “Alice could have known something false” is true in C at $w$ if “Alice knows something false” is true in C at $w^*$. So “Alice could have known something false” is true in C at $w$. By standard semantics, (6.18) is true in C at $w$:

(6.18) Somebody could have known something false.

Or, in a more cumbersome form, “it could have been the case that somebody knows something false”. Now a speaker could utter (6.18) without attending to Alice’s case in $w^*$ and without ceasing to presuppose that $w^*$ does not obtain. So failures of factivity can be uttered. The problem generalises. (6.18) is true in any context such that (a) there is some world $w^*$
6.6. The unmentioned subjects problem

which is properly ignored but satisfies “could have obtained” in C, (b) in that world, some subject has evidence that eliminates \( \neg p \) in the properly ignored worlds, and (c) \( p \) is false in the world of C. An analogous problem arises with “could p” claims, so that “somebody could know something false” is true in C as long as some world \( w^* \) satisfies “could obtain” in C satisfies conditions (b) and (c).

Lewis’s Rule of Actuality was meant to ensure that only what is true is “known” (554):

**Rule of Actuality** For any context C, if a case \( \alpha \) is under discussion in C, \( \alpha \) cannot be ignored in C.

The rule is intended to require that error must be avoided at the actuality of a case for the subject to know in that case. But it guarantees only that the actuality of cases under discussion is relevant to satisfying “know”. The actuality of unmentioned subjects can be ignored, and as a result, it does not follow that only what is true is “known” by unmentioned subjects. Of course the problem is avoided if we say instead that the actuality of any possible case cannot be ignored. But then almost no case could be ignored in any context, contrary to the intention of Lewis and those who follow him.49

6.6.3 Diagnosis: Lewisean vs. DeRose’s semantics

The source of the problem is with the very form of Lewis’s semantics. Lewis’s Rules pick up a set of worlds as a function of context; call them the privileged worlds. One satisfies “know” in that context iff one avoids error at the privileged worlds, *whichever world one is in*. Recall the method version of the semantics, for instance (p. 304):

49. For these reasons it is not the case that “together, the Rule of Actuality and the Rule of Resemblance together function as a safety condition on the satisfaction of ‘know’ ” (Blome-Tillmann, 2009b, 259). The two rules only imply that in any given context, subjects under discussion must satisfy a safety condition in order to satisfy “know”, but the bets are off for unmentioned subjects.

50. von Fintel and Heim (ms, sec. 3.2.3) expose an analogous problem for a candidate context-sensitive semantics for modals. The problem affects Blome-Tillman’s (2009b) and (as far as I can tell) Ichikawa’s (ms.b) variants of the Lewisean semantics.
Generic methods contextualism, Lewis-style For all \( C, \alpha, K_C\alpha \) iff for some method \( m: Bm\alpha \) and for all \( \beta \) such that \( \alpha \in P_C, \neg E_m\beta \).

where \( P_C \) is the union of possibilities that are not properly ignored in \( C \).

Relative to a given context \( C \), the relevant error cases are drawn from a fixed set, \( P_C \). The set is the same for whichever case \( \alpha \) we evaluate. But that is wrong, since the case \( \alpha \) itself may be out of \( P_C \), and so we may end up ascribing “knowledge” in \( \alpha \) despite error in \( \alpha \). On the other hand, we cannot require that all cases \( \alpha \) be included in \( P_C \), since this will make \( P_C \) irremediably sceptical.

What we need instead is to determine a set of relevant alternatives as a function of a case we evaluate. Thus context should not directly provide a set of cases, but an accessibility relation. All the cases that are “accessible” from a case \( \alpha \) are those that are relevant to evaluate knowledge in \( \alpha \). This is in effect what DeRose’s semantics in terms of “strength of epistemic position” or “distance” provides. Recall (p. 304):

Generic methods contextualism, DeRose-style For all \( C, \alpha, K_C\alpha \) iff for some method \( m: Bm\alpha \) and for all \( \beta \) such that \( \alpha R_C\beta, \neg E_m\beta \).

where \( \alpha R_C\beta \) iff the distance between \( w_\alpha \) and \( w_\beta \) satisfies the threshold of \( C: d(w_\alpha, w_\beta) \leq k_C \).

Here the context \( C \) provides a relation \( R_C \) such that to satisfy “know” in \( \alpha \), one should avoid error in cases to which \( \alpha \) bears \( R_C \). So each case gets its own relevant alternative set. If we assume that for each context \( C \), \( R_C \) is reflexive, no case of error can satisfy “know” in \( C \). So factivity is guaranteed. 51

---

51. Revision note. François Récanati asked whether one could defend a version of Lewis’ semantics by introducing MacFarlane’s (2003, 322) distinction between context of utterance and context of assessment (or evaluation). The suggestion is that a knowledge attribution made in a certain context of utterance is true relative to a certain context of assessment iff the subject’s evidence eliminates all the worlds relevant in the context of assessment.

I doubt so, for the unmentioned subject problem recurs in a slightly different form. Either all worlds are relevant in the context of assessment, or they are not. If they are, almost no knowledge attribution will come out true with respect to that context of assessment, for one will almost always find a world in which the subject has the
6.6. The unmentioned subjects problem

6.6.4 A puzzle for DeRose’s Rule of Sensitivity

With the distinction between Lewis’s and DeRose’s semantics clearly in view, a puzzle appears in the application of DeRose’s Rule of Sensitivity. Recall (p. 292):

**Rule of Sensitivity** When a subject’s “knowledge” in \( \alpha \) is under discussion in \( C \), the threshold of \( C \) is normally at least as high as to imply that \( \alpha \) is case of “knowledge” only if it is a case of sensitive belief.

(Where \( \alpha \) is a case of sensitive belief iff \( \alpha \) is a case of belief, and at the closest worlds \( w_\beta \) to \( w_\alpha \) in which \( p_\alpha \) is false, \( S_\alpha \) does not believe \( p_\alpha \).)\(^{52}\)

First, consider the simplest cases, in which “knowledge” is attributed to an actual subject. (That is, the world of the context and the world of the subject coincide.) Focus, for instance, on Bouvard’s ascriptions of knowledge to the visitor (6.13, slightly amended), and assume that the Rule is applied throughout. First, Bouvard utters:

\[(6.19) \text{She knows that the animal in the pen is a zebra.}\]

This is context \( C_1 \) and case \( \alpha \). By the Rule of Sensitivity, since \( \alpha \) is under discussion, sensitive belief in \( \alpha \) is required for \( \alpha \) to satisfy “know” in \( C_1 \). This implies that error should be avoided as far as the closest cases in which the animal in the pen is not a zebra. These are not far away: there are the worlds in which the antelope’s pen and the zebra’s pen have been same evidence (in Lewis’s sense) and the target proposition is false. If they are not, then it will typically be possible to find a subject located in the non-relevant worlds, such that the evidence of the subject in that world is incompatible with the falsity of the target proposition at the relevant worlds, even though the proposition is false at that irrelevant world. In the latter case we have a failure of factivity relative to the context of assessment.

\(^{52}\) For reference, here is DeRose’s (1995, 36) formulation: “When it is asserted that some subject \( S \) knows (or does not know) some proposition \( P \), the standards of knowledge (the standards for how good an epistemic position one must be in to count as knowing) tend to be raised, if need be, to such a level as to require \( S \)’s belief in that particular \( P \) to be sensitive for it to count as knowledge.” My gloss is: “When it is asserted that some subject \( S \) knows \( P \) in some case, the standard tends to be raised so as to require \( S \)’s belief in \( P \) to be sensitive in that case.” The precision, which is crucial for the problem raised below, is needed when the case under discussion is not an actual one.
exchanged, for instance. So $k_{c_1}$, the modal distance at which error should be avoided, is low. Now Pécuchet utters:

(6.20) Well, it could have been a painted mule! Does she know whether it is not a mule cleverly painted to look like a zebra?

This is context $c_2$, and case $\alpha'$. (Case $\alpha'$ involves the same subject, world and time as $\alpha$, but a different proposition.) By the Rule of Sensitivity, since $\alpha'$ is now under discussion, sensitive belief in $\alpha'$ is required for $\alpha'$ to satisfy “know” in $c_2$. Error should be avoided as far as the closest cases in which the animal is a painted mule. Now these are, by assumption, further away. So $k_{c_2}$ is greater than $k_{c_1}$. Assuming also that at the closest mule world, the visitor does believe that the animal is not a mule, $\alpha'$ does not satisfy “know”. But the new standard $k_{c_2}$ applies to all cases evaluated in $c_2$, and since there is a close possibility within the $k_{c_2}$ distance in which the visitor wrongly believe that the animal is a zebra (namely, the mule possibility), $\alpha$ also fails to satisfy “know” in $k_{c_2}$. So:

(6.21) She does not know that the animal in the pen is not a cleverly painted mule.

(6.22) She does not know that the animal in the pen is a zebra.

are both true in $c_2$.

But now consider a variant of the case, in which Pécuchet does not utter (6.20) but the following:

(6.23) Well, it could have been a painted mule! If it had been a painted mule, she would have believed that it was a zebra all the same.

In this context $c_3$, Pécuchet is bringing forward a third case, $\alpha''$, in which the visitor looks at a painted mule and believes that it is not one. By the Rule of Sensitivity, error should be avoided as far as the closest cases in which the animal is a painted mule. But closest to whom? Since the idea is that sensitive belief is required in the case under discussion, $\alpha''$, we require that error should be avoided at least as far as the distance that separates $\alpha''$ from its closest mule world — not us, speakers, from our closest mule world. But $\alpha''$ is a mule case. Requiring sensitive belief in $\alpha''$ that the
animal is not a mule is just requiring one not to believe in \( a'' \) that the animal is not a mule. So there is no need to raise the standards, \( k_{C_3} = k_{C_1} \). If Bouvard now reasserts (6.19), his utterance is true.

Of course if Bouvard now asserted “She knows that the animal is not a painted mule” — or “She does not know that the animal is not a painted mule”, for that matter —, he would thereby introduce a case of belief in the actual world that the animal is not a painted mule. The resulting context would be as in \( C_2 \), and an assertion of (6.19) would come out false. But supposing that he does not, Pécuchet’s utterance of (6.23) by itself has failed to raise the standards. This is entirely unexpected on DeRose’s outlook, since (6.23) is an utterance of the sensitivity conditional which DeRose (1995, 17–20) regards as the paradigm case of a claim that sways our intuitions on the sceptical side.

One reply that DeRose could make is to say that Pécuchet’s assertion implicitly brings under discussion the case of the visitor’s actual belief that the animal is not a mule. The standards are also raised to require sensitive belief in the implicitly discussed case, and so they go up as in \( C_2 \). I doubt that the reply will be sufficiently general. Call the current visitor “Alice”. Suppose Pécuchet asserted instead:

(6.24) Yesterday another visitor [Bob] was looking at a painted mule that looked just like that, and he took it for a zebra.”

The utterance is as efficient as the others to press the intuition that Alice does not know that the animal is a zebra. But the case explicitly brought into focus, Bob’s, is a case of false belief, so applying the Rule to his case does not raise the standards. Are we implicitly discussing Alice’s current belief that yesterday’s animal was not a mule? She might not even have that belief. If she had that belief, it would be a false one, and again there is no need to raise the standards. The only way to raise them in that case is to assume that we implicitly discuss Alice’s current belief that today’s animal is not a mule. But how or why this is done is not clear.

Note finally that DeRose could not amend the rule by saying that when \( P \) is under discussion, the standards are raised, if need be, such
that “knowledge” requires sensitive belief that P for any belief that P in any case. Take just about any ordinary proposition that is commonly said to be “known” by some subjects, such as the proposition that it is sunny over London now. There will be a (typically unmentioned and counterfactual) subject such that to require sensitive belief of him for that proposition will be excessively demanding — say, a subject in a world in which the weather is constantly grey in London as a matter of physical necessity. Standards for “know” would be raised very high, and most ordinary knowledge claims would turn out false.

6.7 Conclusion

We have examined Lewis’s and DeRose’s views according to which the semantic value of “know” is affected by a contextually-variable set of relevant alternative cases in which error should be avoided. We have pointed out that both postulate context-sensitivity rules that are specific to “know”. We have raised difficulties for both proposals.

In Lewis’s case, the context-sensitivity mechanism is simple and easily motivated. Once a possibility is attended to, or taken seriously, it is included in a set of relevant possibilities. “Knowledge” then refers, in context, to the absence of error in that set. The idea faces two major problems. First, having a semantics in terms of a contextually selected set of possibilities, rather than an accessibility relation, leads to an unacceptable violation of factivity (6.6.2). Second, the idea cannot account for the appeal of third-person sceptical arguments (6.5.3). The notion of attending to a possibility, or not presupposing a possibility, also raises some worries (6.2.4).

DeRose’s semantics is formulated in terms of strength of epistemic position, to which correspond accessibility relations. The factivity problem is avoided. But it is much less clear how the context-sensitivity rule that DeRose posits works. We have pointed out that it does not seem to have any clear analogue elsewhere in language (6.3.5), and that its application is unclear in the paradigmatic case of someone uttering a sensitivity
condition (6.6.4).
Chapter 7

Modals and knowledge

In the present chapter I formulate an alternative contextualist semantics for knowledge claims. On modal infallibilist accounts of knowledge, knowledge is centrally a matter of impossibility of error. Both DeRose and Lewis think that this fact is reflected in the semantics of ordinary knowledge claims. Speakers do implicitly represent sets of relevant alternative possibilities and are sensitive to facts about such sets in their ascriptions or denials of “knowledge”. I think the idea is right, but the execution flawed. Both DeRose and Lewis postulate sets of possibilities and context-sensitivity rules that are specific to “know”, and each runs into a number of difficulties, as we have seen.

Section 7.1 lays out the view of modals we work with (Kratzer, 2010b; Hacquard, 2006) and isolate the “circumstantial” interpretation of modals that interests us. Sections 7.2–7.3 argue for the existence of a specific contextual parameter, real possibility, that is at play in the semantics of circumstantial modals and subjunctive conditionals. Section 7.4 uses the notion to provide an alternative modal contextualist account, according to which “knowledge” requires a method that avoids real possibilities of error. The real possibility account avoids a number of problems that Lewis’s and DeRose’s accounts faced, and predicts a number of links between knowledge ascriptions and subjunctive conditionals. In Section 7.5, I discuss whether we should endorse the account. The matter is
difficult, and I leave it undecided. But I submit that contextualism about circumstantial modals and about knowledge stand or fall together.

7.1 The semantics of modals

We first give an overview of modal auxiliaries based on Kratzer (1981, 2010b) and Hacquard (2006). This allows us to single out the ones I am interested in: “circumstantial” modals, that is, modals expressing an alethic modality.

7.1.1 Kratzer’s semantics for modals

Modal auxiliaries like must, may, might, can, could in English can express a wide variety of meanings:

(7.1) (a) Greg must be home. [Since he is not at work or at the bar.]
(b) Greg must stay home. [He has to watch over the kids.]
(c) Greg must stay home. [If he doesn’t want to get sick.]
(d) Greg must stay home. [Since he is not able to walk.]

(7.1 a) expresses what must be the case given what we know, (7.1 b) what Greg must do given his duties, (7.1 c) what Greg must do if certain aims are to be reached, (7.1 d) what will happen given Greg’s (in)abilities. The same variability is found cross-linguistically. As Kratzer (1977) emphasised, postulating ambiguity is implausible. Instead, she considers modals as quantifiers over contextually specified domains of possibilities. The domain of possible worlds that are compatible with what the speaker knows yields an epistemic reading, the domain of possible worlds in which one’s duties are obeyed yields a deontic reading, and so on. This view of modal constructions is now widely accepted.1

1. Though not universally. A number of authors argue that epistemic modals do not contribute to truth conditions, but are rather evidential markers or speech act modifiers. See Papafragou (2006) and Portner (2009, 145–151) for an overview and discussion. The agreement is wider on the kind of (use of) modals we are interested in, however, namely alethic ones.
7.1. The semantics of modals

*Kratzer* (1981, 2010b) proposes the following general semantics for any modal MOD:

**Kratzer’s semantics**  
MOD p is true in C at w iff Q of the g(w)-best f(w)-worlds are worlds in which p is true in C.

The semantics has three parameters.

1. The quantificational force Q. Must p is true iff all worlds in a certain range are p worlds; could p is true iff some world in a certain range is a p world. In English, the quantificational force is encoded in the lexical entry of the modal: the necessity modals must, have to, need are always universal quantifiers over possibilities, while the possibility modals can, may, might are always existential quantifiers.

2. The modal base f. The modal base provides a first restriction of the domain of possible worlds the modal quantifies over. For instance, for an epistemic reading, f provides at each world w a set of possible worlds that are compatible with the evidence available in w. Thus “Greg must be home” is true at w iff “Greg is home” is true at all worlds w’ such that the evidence in w’ is the same as the evidence in w. (In Kratzer’s semantics, f is a function from worlds to sets of propositions. The corresponding set of worlds is \( \cap f(w) \), the set of worlds w’ in which all the propositions in f(w) hold.)

3. The ordering source g. Kratzer introduces a separate restriction on the domains to handle sub-optimal deontic claims, graded modalities and subjunctive conditionals. Consider a claim such as: “Lucy must go to jail”. This cannot mean that in all the worlds where the law is obeyed, Lucy goes to jail, since in those worlds, Lucy has not committed a crime and should not be put to jail. Rather, it means that within the worlds in which the current circumstances hold (namely, Lucy has committed a crime), the most lawful ones are ones in which she goes to jail. We thus have an ordering given by g, such that at each world w, and for any restricted set of worlds, the ordering selects a set of g(w)-best worlds: the worlds that obey most of the laws of w, those that satisfy most of the desires of a
relevant agent in \( w \), those that exhibit most of the regularities and stereotypical features of \( w \), those that are most similar to \( w \), and so on. (In Kratzer’s semantics, \( g \) is a function from worlds to sets of propositions. A world \( x \) is at least as good as a world \( y \) by the lights of \( g(w) \) iff all propositions in \( g(w) \) that are true in \( y \) are also true in \( x \): \( x \leq_{g(w)} y \) iff \( \{ p : p \in g(w) \land p \text{ holds in } y \} \subseteq \{ p : p \in g(w) \land p \text{ holds in } x \} \).

Formally, let \( f \) and \( g \) be functions from worlds to sets of propositions, and let propositions be sets of worlds. The Limit Assumption is the claim that \( g(w) \) is such that for any set \( S \) of worlds, there is a set of best \( g(w) \)-worlds in \( S \) — that is, we do not have a infinite sequence of increasingly better subsets of \( S \). 2 Granting the Limit Assumption, we can state:

**Simple Kratzerian semantics**

\[
[\text{MOD} \, p] = \{ w : (Q : w' \in S(w))(w' \in [p]) \},
\]

where

for any \( w \), \( S(w) = \{ w' : w' \in f(w) \land \forall w'' \in f(w)(w' \leq_{g(w)} w'') \} \),

for any \( w, w', w'' \), \( w' \leq_{g(w)} w'' \) iff \( \{ q : q \in g(w) \land w' \in q \} \subseteq \{ q : q \in g(w) \land w'' \in q \} \).

That is, \( \text{MOD} p \) is true at a world \( w \) in a context \( c \) iff \( Q \) of the \( S(w) \) worlds are worlds at which \( p \) is true in context \( c \). Where the Limit Assumption fails, \( S(w) \) is empty, and the simple semantics is trivialised: all necessity claims true and all possibility claims false. To avoid making the Limit Assumption, a more complex semantics is needed, but the complication can be ignored for our purposes. 3

---

2. See Lewis (1973, 18–21).

3. Here is the complication. Instead of taking the best sphere, we say that the modal claim holds iff the corresponding modal property holds in *some sphere and all the better spheres*. Write \( S(w, w') \) for the sphere of \( f(w) \)-worlds that are at least as \( g(w) \)-good as \( w' \), for any \( w, w' : S(w, w') = \{ w'' : w'' \in f(w) \land w'' \leq_{g(w)} w' \} \). Now we say that \( \text{MOD} p \) holds in \( w \) iff there is a \( w' \) such that for all \( w'' \) s.th. \( w'' \leq_{g(w)} w', (Q : w'' \in S(w, w'))(p \in w') \). That is, \( Q \) of the \( S(w, w') \) worlds are \( p \) worlds, and for any better sphere \( S(w, w'') \), \( Q \) of the \( S(w, w'') \) are \( p \) worlds. The semantics for necessity and possibility modals in Kratzer (2010b, 18–9) is derived by taking \( Q \) as a universal and existential quantifier, respectively.
7.1. The semantics of modals

In English, modal base and ordering source are provided by context.\(^4\) Kratzer (2010b, 12–5) distinguishes three types of modal bases:

1. Circumstantial: for each \(w\), \(f(w)\) is a relevant set of facts of \(w\).

2. Evidential: for each \(w\), \(f(w)\) is a relevant set of known facts (a body of evidence) of \(w\).

3. Informational: for each \(w\), \(f(w)\) is the content of some source of information in \(w\).

Circumstantial and evidential bases both state a relevant set of facts in each world. They are both “realistic”: each proposition in \(f(w)\) is true in \(w\). They only differ in how the facts are picked up.

Informational bases are a recent addition to Kratzer’s views — they were absent from Kratzer (1981) but introduced in its recent revision (2010b, 12–5, 33–4). They need not be “realistic”: if the source of information is unreliable, some propositions in \(f(w)\) may be false in \(w\). Kratzer illustrates the difference between evidential and informational bases with the following pair:

\[(7.2)\]

(a) Given the rumour, Roger must have been elected chief.

(b) According to the rumour, Roger must have been elected chief.

\(7.2a\) states that it is incompatible with the existence of the rumour that Roger has not been elected; it follows that he has been elected. \(7.2b\) says instead that it is incompatible with the content of the rumour, and can be uttered by someone who thinks or knows that he has not been elected. However, Kratzer (2010b, 13n) hesitates between counting informational content as a modal base or ordering source.

Kratzer (2010b, 15–6) distinguishes three types of orderings:

1. Stereotypical: for each \(w\), \(g(w)\) states what is normal in \(w\), by some relevant standard of normalcy.

---

\(^4\) These features are widespread cross-linguistically but not universal. In a now famous study, Rullman et al. (2008) show that the St’át’imcets language (Lillooet Salish, a language with a few remaining speakers in British Columbia, Canada) lacks the common contrast between possibility and necessity modals. They argue that in opposition to familiar Indo-european languages, St’át’imcets modals lexically encode their modal flavour (e.g., epistemic or deontic) but not their quantificational force (possibility vs. necessity), which is instead provided by context.
2. Deontic: for each $w$, $g(w)$ states the content of relevant laws, duties, norms, aims, desires or preferences in $w$.

3. Totally realistic: for each $w$, $g(w)$ is a full description of $w$, that is, a set of propositions that characterises $w$ uniquely.

Kratzer (2010b, 45) uses totally realistic ordering sources to provide a Lewis-like semantics of subjunctive conditionals (Lewis, 1973). A totally realistic ordering source in effect orders worlds according to how much of the description of $w$ they satisfy. The “closest” worlds are worlds in which most of the propositions that hold in $w$ also hold.

Deontic orderings rank possible worlds according to how much of our laws, duties, aims or desires are obeyed or satisfied in those worlds. Consider for instance:

(7.3) To see real zebras, you must go to the zoo.

(7.3) expresses a teleological modality: what must be done in order to reach a given aim. On Kratzer’s semantics, its meaning is derived with a circumstantial modal base and a teleological ordering. We only consider worlds in which certain circumstances of the actual world are fixed, such as the location of zebras, the available means of transportation, and so on. Those worlds are then ordered according to whether they satisfy the aim of (the addressee’s) seeing real zebras; the best ones are those in which she does. If in all the best ones, the addressee goes to the zoo, (7.3) is true. A similar story holds for other deontic or normative modal claims (Kratzer, 2010b, 34–41). Normative orderings need not count the actual world among the best ones; thus “Roberto must do the dishes” may be true at a world in which Roberto does not do the dishes.

Kratzer uses stereotypical orderings to account for graded modalities, as in:

(7.4) Nick might very well be at the bar.

The idea is that for (7.4) to be true, it is not sufficient that there is a possible world compatible with our evidence in which Nick is at the bar.

---

5. The idea is that the phrase “To see real zebras” somehow binds the ordering parameter $g$ of “must”.
The world(s) in question should not be exceptional or far-fetched at all. It has to be a normal or stereotypical possibility that Nick is at the bar. Kratzer implements the idea by using a function $g(w)$ that yields a set of propositions stating what is normal or stereotypical in $w$. (The set need not be consistent, and $w$ need not be perfectly normal by its own standards, as with normative orderings.) Worlds are then ordered with respect to how much of the stereotypes of $w$ they satisfy. In appropriate contexts, a modal such as “might very well” quantifies over the best — that is, most typical or normal — worlds in a given range.\footnote{Since on Kratzer’s view, the actual world need not be the best one according to its own normalcy standards, her semantics imply that “might well $p$” claims can be false even when $p$ is true — assuming that “might well” has a stereotypical-ordering semantics. For instance, “Nick might very well be at the bar” can be false even if Nick is in fact at the bar. My intuitions are not clear on such cases.}

It is tempting to amend Kratzer’s proposals by leaving out informational sources and the ordering specific to subjunctive conditionals. The resulting picture draws an attractively simple division of labour between bases and orderings:

1. **Modal bases are realistic.** They are either circumstantial or evidential.
   They restrict to a fixed set of circumstances or a fixed body of evidence. Realistic bases ensure that the set of accessible worlds is nowhere empty: for each $w$, $w \in \bigcap f(w)$.

2. **Ordering sources order accessible worlds by some ideal, such as normalcy, duty, desires, goals and so on.** Ordering sources need not be realistic: the actual world may turn out not to be moral, pleasurable or normal. Ordering sources need not even be consistent (see Kratzer, 2010c, 15–21).

The semantics of subjunctive conditionals we will use does not rely on a Lewis-style ordering, so we do not need totally realistic ordering sources. Whether we can do without informational modal bases requires further discussion.\footnote{One problem with taking informational sources as modal bases is that informational sources may be inconsistent. Inconsistent modal bases trivialise the semantics, so some tinkering would be needed. An alternative option is to take informational...} At any rate, nothing I defend here depends on the simplifi-
cation. (If a realistic ordering is needed for some other purpose than the semantics of counterfactual, I am happy to accept it.)

7.1.2 Hacquard on root vs. high modals

Kratzer’s semantics provides a unified account of modal auxiliaries and of the diversity of their readings. However, it allows more diversity than is actually observed. Modal readings appear to be constrained in a number of ways — we will see some examples below. Hacquard (2006, ch. 3) has proposed to derive the constraints from the different syntactic positions modal auxiliaries may take. The overall picture Hacquard reaches will allow me to single out the modal readings that interest us here.

The following French sentence is ambiguous between two readings:

\[(7.5) \text{Jane a dû prendre le train.}\]

Jane must-past-perfect take the train.

content as an ordering source, as Kratzer (2010b, 13n) suggests. But they should then be paired with an empty base in order to allow for “counter-circumstantial” information (“Roger could not be in the Bahamas, since he is dead, but according to the rumour, he must be there.”). This predicts (rightly?) the impossibility of analogues of sub-optimal deontics (such as “She must go to jail”) with informational readings. Also, as Kratzer (2010b, 13n) notes, stereotypical orderings may have to be piled with the informational ones, which complicates the semantics. Pairing with normative orderings would be needed as well, as shown by “according to the rumour, Roger has to do the dishes [in order to be elected].” Another option is to eliminate informational readings altogether. Kratzer puts forward two illustrations in English. The first is the example (7.2b above:

“According to the rumour, Roger must have been elected chief” (2010b, 14, 33). But the “According to . . . ” phrase can be seen as a report device, as in “according to Sally, Roger is chief”. Its semantics would roughly be: in every world \(w\) compatible with the content of Sally’s beliefs, Roger is chief. “According to Sally, Roger must be chief” would then mean: for each world \(w\) compatible with Sally’s beliefs, in all worlds \(w’\) compatible with the evidence of \(w\), Roger is chief. In other terms: Sally believes that we are in a world in which the evidence indicates that Roger is chief. Thus we do not have an informational modal, but an evidential one embedded in an attitude operator. The second is a case in which the speakers are looking at a cabinet containing evidence about a murder, and bet on who is the murderer according to the cabinet. They utter “Kastenjakl must have done it” and “Gauzner-Michl might have done it” (2010b, 33–4). But the case can be seen as one of implicit report, in effect equivalent to “According to the cabinet, . . . ”. The same reply is thus available. Better evidence for Kratzer’s informational readings may come from the reportative sollen in German and the reportative conditionnel in French.
Epistemic: Given my evidence now, it must be the case that Jane took the train then.

(Goal-oriented) Circumstantial: Given Jane’s circumstances then, she had to take the train then (to reach her destination). (Hacquard, 2006, 25)

On the first reading, the necessity modal *devoir* is keyed to *features of the discourse situation*, namely the evidence of the speaker at the time of utterance. On the second reading, the modal is keyed to *features of the topical situation*, namely the circumstances Jane was in at the time we are talking about. We cannot get a reading which is keyed to the speaker past evidence, for instance, even though such a reading would be available on Kratzer’s semantics.

Hacquard (2006, ch.3, §4, esp. 134–5) derives the ambiguity from the position of the modal auxiliaries in the syntactic tree. The modal can be placed above or below Tense. Roughly, the two possible trees are:

1. Must [ Past [ Jane takes the train ] ]
2. Past [ Must [ Jane takes the train ] ]

Hacquard argues that when the modal is high, as in (1), its modal base is keyed to features of the speech situation. When the modal is low, as in (2), its modal base is keyed to features of the topical situation. This explains why the combinations *speaker’s evidence* + *present time* and *subject circumstances* + *past time* are the only ones available.

When a modal is high but embedded in an attitude verb, it is keyed to features of the attitude, instead of features of the speech context:

(7.6) Serge pense que Jane a dû prendre le train.

Serge think-present that Jane must-past-perfect take the train.

Epistemic: Serge thinks that given his evidence now, it must be the case that Jane has taken the train.

(This latter aspect of Hacquard’s account is problematic in view of the ease with which we disquote epistemic modals. I will not get into that

8. More precisely, above or below the combination of Tense and Aspect.)
issue here. 9)

Modal auxiliaries are thus split into two categories, according to whether they are placed above or below Tense. *High* modals are keyed to the speech act situation (in the non-embedded case); they are *speaker/addressee oriented*. They divide into *epistemics* and *true deontics*. Epistemics state what must or may be the case given evidence available in the speech situation. True deontics state what must or may be the case given norms (duties, goals, . . . ) in place in the speech situation. For instance, they may express an obligation put on the addressee:

(7.7) Lydia must brush her teeth. [said to Lydia’s babysitter by Lydia’s mother.] (Hacquard, 2006, 124)

Or an obligation endorsed by the speaker:

(7.8) I must become a baker. (see Kratzer, 2010b, 39)

By contrast, *low* modals — called *root* modals in the literature — are keyed to the topical situation, the one relevant for the verb below the modal: they are *subject-oriented*. They further divide into *pure circumstancials* and *goal-oriented circumstancials*. Pure circumstancials state what has to or can happen given relevant circumstances of the topical situation. Goal-oriented circumstancials state what has to or can be done in relevant circumstances to reach some goals of the subject. Goal-oriented circumstancials are a species of deontic modality, but the relevant norms are now those in place in the subject situation:

(7.9) Jane had to take the train.

(See Hacquard, 2006, 40–43 for further discussion of subject-oriented deontics.)

Summing up, we get four types of interpretations modal auxiliaries (Table 7.1):

Hacquard (2006, 136–50) associates the interpretations with three accessibility functions. On her semantics, sets of possible worlds are given as a function of events, not of worlds as in Kratzer’s semantics. Note that

7.1. The semantics of modals

<table>
<thead>
<tr>
<th></th>
<th>factual</th>
<th>normative</th>
</tr>
</thead>
<tbody>
<tr>
<td>speaker/ addressee oriented</td>
<td>epistemics</td>
<td>true deontics</td>
</tr>
<tr>
<td>subject oriented</td>
<td>circumstantial</td>
<td>subject-oriented deontics</td>
</tr>
</tbody>
</table>

Table 7.1: Four classes of modal auxiliaries.

Events are not world-bound: the same event typically occurs at several worlds. But they have many of their properties essentially: an actual slapping of Guy by Gill is an event of Gill slapping Guy at every world in which it exists. The accessibility functions map each particular event to a set of worlds. Hacquard (2006, 150) gives three:

1. Epistemic. \( f_{\text{EPISTEMIC}}(e) = \{w: w \text{ is compatible with the content of } e\} \).
2. Circumstantial. \( f_{\text{CIRCUMSTANTIAL}}(e) = \{w: w \text{ is compatible with the circumstances of } e\} \).
3. Deontic. \( f_{\text{DEONTIC}}(e) = \{w: w \text{ is compatible with the To-Do list of the addressee of } e\} \).

These are three particular functions, not types of functions. Hacquard assumes a function that gives us, for any event \( e \), the content of \( e \), if any. When \( e \) is the speech event, its content is the set of propositions believed by the speaker (141–5). When \( e \) is a propositional attitude event, we get the contents of the attitude: all the propositions believed, for instance (136–41). Furthermore, for any event \( e \), Hacquard assumes a function that gives the circumstances of \( e \): a set of relevant facts about \( e \), such as its participants, location, time (144–7). Finally, for any speech event \( e \), we get a unique set of items on the addressee’s To-Do List (147–50).

10. One may prefer the set of facts known for the content of speech and attitude events, but Hacquard (2006, 144) opts for belief. Only “acceptance” attitudes — those that are correct only if true, such as belief, knowledge, assertion — can provide contents for modals. See Anand and Hacquard (2009) for further discussion.
High modals are bound to the event referred to by the attitude verb above them, if they are embedded under an attitude, and by the speech act event otherwise (Hacquard, 2006, 137–8; 2010, 105–6). Low modals are bound to the event referred to by their complement clause (Hacquard, 2006, 146; 2010, 110). Thus “Nick had to stay home” is true iff there is an event of Nick staying home whose circumstances entail that Nick stayed home. It is not entirely clear to me what goes on when the event referred by the clause is not actual, as in “Nick could have stayed home [but did not]”.  

Hacquard (2010, 110) discusses the case of “Mary had to take the train to Paris [but did not]”. Her idea is that the imperfective tense shifts the evaluation of the whole sentence to a counterfactual world. Thus the sentence is true iff there is a (relevant) counterfactual world in which there is a train-taking by Mary, and the circumstances of this train-taking event are such that in all worlds where they obtain, Mary takes the train to Paris. Alternatively, we could say that when the event is not actualised, the modal is bound to a “stand-in” actual event, which is roughly the shadow in the actual world of the non-actualised event described by the clause — for instance, an event of Nick not staying home could serve as stand-in for the event of Nick staying home. The circumstances of the stand-in event would give us “the circumstances in which the clause event is to take place” (Hacquard, 2006, 145), as desired.

Hacquard’s accessibility functions have two problematic features. First, they do not account for subject-oriented deontics such as (7.5). Such modals require both a restriction to Jane’s past circumstances and her past To-Do list. But Hacquard’s semantics only allow them to pick one or the other. A natural solution to the problem is to reintroduce Kratzer’s orderings. Modals have two parameters: base and ordering. Deontic readings are obtained by a normative ordering, not by a normative base. This suggests the following picture. When the modal is high, its base is always epistemic. When it is low, its base is always circumstantial.

11. I should record Hacquard’s caveat here: “a comprehensive semantics of the event dependence of this modal base will have to await another occasion” (Hacquard, 2010, 109).
7.1. The semantics of modals

In both cases, the ordering may be deontic or not. A true deontic is a deontically ordered epistemic, a subject-oriented deontic is a deontically ordered circumstantial.

Second, Hacquard’s accessibility functions threaten to invalidate some apparently harmless inferences such as:

(7.10) Yesterday, Oscar had to stay home. Yesterday, Pedro had to stay out. So yesterday, they could not meet.

Write $\tau(e)$ for the time of $e$ and $Y$ for the time of yesterday. (Times are sets of instants.) On Hacquard’s semantics, we get:

(7.11) There is a past event $e$ such that $\tau(e) \subseteq Y$ and for all $w$ in $f_{\text{CIRCUMSTANTIAL}}(e)$, Oscar stays at home in $w$.

There is a past event $e'$ such that $\tau(e') \subseteq Y$ and for all $w$ in $f_{\text{CIRCUMSTANTIAL}}(e')$, Pedro stays out in $w$.

So there is a past event $e''$ such that $\tau(e'') \subseteq Y$ and for no $w$ in $f_{\text{CIRCUMSTANTIAL}}(e'')$, Oscar and Pedro meet in $w$.

The inference is not valid because each sentence picks a different event: an event of Oscar staying home, an event of Pedro staying out, and (perhaps) an event of Oscar and Pedro not meeting. The different events have different circumstances, which are likely to yield different possibility sets. So the inference may fail.

There are two options for ensuring that the inference is valid. One is to use big events. We say that all three occurrences of the past pick up a single event. Presumably, it is going to be a big one: the whole of what occurred yesterday, for instance. Another is to use common possibilities. We say that all events that are at the same time share the same circumstances. So there is a single set of circumstantially accessible worlds for all contemporaneous events. Whichever option we choose, the circumstantial accessibility relation for events will mimic an accessibility relation from world and times.

**Circumstantial accessibility is coarse** For all $e, e'$: if $e$ and $e'$ occur in $w$ and $\tau(e) = \tau(e')$ then $f_{\text{CIRCUMSTANTIAL}}(e) = f_{\text{CIRCUMSTANTIAL}}(e')$. 
We define $f_{\text{CIRCUMSTANTIAL}}^{*}(w, t) = f_{\text{CIRCUMSTANTIAL}}(e, \tau(e))$ where $e$ is any event that occurs in $w$ such that $\tau(e) = t$.\footnote{We may want to generalise the circumstantial equivalence to all events which share the same end time.}

The coarseness of circumstantial accessibility makes the event-relativity of circumstantial modals idle. That does not mean that event-relativity should be rejected. It may still play a role with epistemics and deontics. And more importantly for us, it may be that not all circumstantial modals are event-insensitive. I leave it open that some ability readings involve person-relative sets of possibilities.

Summing up, we distinguish high and low modals, following Hacquard. High modals are epistemic, low modals are circumstantial. Both can be given a deontic flavour through a normative ordering; we call the resulting interpretation true deontics and subject-oriented deontics, respectively. While we may need fine-grained event-relative accessibility relations for epistemic and deontics, we need a coarse world-time accessibility relation for some circumstantial interpretations at least.

The modal interpretation we are interested in here is the coarse circumstantial one. In the next two sections, we provide further evidence that there is such an interpretation.

### 7.2 Circumstantial modals and real possibility

In this section, I want to argue for the existence of a specific parameter of contexts of conversation: the accessibility relation for real possibility. The relation is part of the background of the conversation, it is not introduced by an overt or covert expression. In intuitive terms, the relation tells, for any world of evaluation, what could really have happened at that world. Formally, we represent the real possibilities of a given world as a set of worlds.

There are two arguments in favour of real possibility. First, circumstantial modals are much more resistant to context-sensitivity than has been assumed. The reluctance is explained on the assumption that all
operate on a single set within a single context. Second, such a set is arguably at play in the semantics of subjunctive conditionals (von Fintel, 2001, Gillies, 2007). The present section gives the first argument, the next one the second.

7.2.1 Disagreement cases

Consider the following case.

**Cannes Trip** Delphine and Solange took their car for a trip to Cannes on a whim. But their car is in a very bad condition. The motor gets too hot and breaks down on the way. They have the following dialogue:

(7.12) *Delphine.* What a shame. We could have reached Cannes!
*Solange.* No way! With the motor in that shape, we couldn’t have reached it.
*Delphine.* But the motor could have been in a better shape. If it had been, we would have reached Cannes. So we could have reached Cannes!
*Solange.* Well if the motor had been in a better shape, we could have reached Cannes. But it wasn’t. So we couldn’t have reached it.

I take it that the dialogue is natural enough. The modals are clearly circumstantial: the travellers know that they have not reached Cannes, and they are not discussing whether they were allowed to reach it. It is also quite clear what is roughly going on. Solange is considering what could have been the case given the actual state of the car, while Delphine is also considering what could have been the case given some other state of the car. We may picture the difference on a tree of branching time. Let \( t \) be the time at which it was settled that the motor would be in bad shape — for instance, \( t \) is the last time when they considered bringing the car to a routine check. Solange is considering the histories that were open from a time after \( t \), while Delphine is considering the histories that were open from some time before \( t \). If Delphine is right, then in the latter set of
possible histories, there is a world in which they reach Cannes. If Solange is right, then in the former, there is none. And they can be both right.

On a straightforward contextualist account, we would associate each occurrence of a modal with its own accessibility relation. In Delphine’s utterances, “could” is associated with an accessibility relation that gives all the possible histories open from a certain time before $t$. In Solange’s utterances, “could” is associated with a different accessibility relation that only gives the possible histories open from the relevant time after $t$. If each of them is right about what goes on in the possibilities relevant to her claim, then a world in which they reach Cannes is accessible by the first relation but not by the second. Thus Delphine’s utterance of (7.13) comes out true:

(7.13) We could have reached Cannes.

And Solange’s utterance of (7.14):

(7.14) We couldn’t have reached Cannes.

comes out true as well.

The simple contextualist treatment is at odds with the speakers’s persisting disagreement, however. The case is arguably a paradigmatic case of use of circumstantial modals. The speakers are making their grounds for judgements quite explicit. If the simple contextualist account is right, both of their utterances can be true. Yet they persist in their attempts to settle the question whether they “could have” reached Cannes or not. Why? The most plausible answer is that each believes that if her utterance is true, the other’s utterance is false. For instance, Solange has a belief that she would express with (7.15).

(7.15) Delphine’s utterance of “we could have reached Cannes” is true iff we could have reached Cannes.

13. See sec. 6.1 on the notion of paradigmatic case of application.

14. To the sceptical reader who would doubt the plausibility of a persisting disagreement of this kind, I suggest an easy experiment: ask the spectators of some sports event in a pub whether the team that lost could have won.
On the simple semantic account, this metalinguistic belief is false.\textsuperscript{15} Solange and Delphine are mistaken about some aspect of the semantics of their claims. Not entirely mistaken, since Solange correctly believes, for instance, something that she would express by:

(7.16) My utterance of “we could have reached Cannes” is true iff we could have reached Cannes.

This is not trivial, even though Solange is in a position in which it is both extremely easy and uninformative to realise it. But she is nevertheless mistaken about an important aspect of the semantics of “could” claims. In the literature on epistemic contextualism, this idea is known as the idea that speakers are semantically blind (Hawthorne, 2004, 107–8).\textsuperscript{16}

We propose instead a semantics on which circumstantial modals cannot take multiple values in a single stretch of conversation. (This is what DeRose (2004, 6) calls a single scoreboard semantics.) A context normally provides a unique accessibility relation for circumstantial “could”, or a unique set of such relations if there is indeterminacy.\textsuperscript{17}

The view is not entirely immune to semantic blindness. First, semantic blindness may result from indeterminacy. If there is indeterminacy, Delphine may have a mistaken belief that she would express by:

\textsuperscript{15} The belief is an instance of a disquotation principle (see Hawthorne, 2004, 101). Disquotation of context-sensitive expressions across relevantly different typically fails.

\textsuperscript{16} See Schiffer (1996, 326) and Williamson (2005a, 223–4). DeRose has argued for the existence of semantic blindness with “know” (1992, 927, see DeRose, 2006 for fuller discussion). Note however that DeRose does not endorse the simple contextualist account of disagreement (DeRose, 2004, 5–7).

\textsuperscript{17} Hacquard’s stated semantics for circumstantial (2010, 109–10) implies this. The event picked up by their clause is that of Solange and Delphine reaching Cannes. (Or, in the “stand-in” approach sketched sec. 7.1.2, a stand-in event of them not reaching Cannes.) There is a single accessibility function that determines the circumstances of that event. If the circumstances entail that they did not reach it, Solange is right, if not, Delphine is. More generally, her semantics implies that circumstantial modals are not context-sensitive. The set of possibilities they rely on can vary with the complement cause of the modal, since the accessibility relation is bound to the event described by that clause. But it cannot vary with context. However, Hacquard (2006, 146) grants that there is vagueness and context-sensitivity in determining the circumstance of an event. This suggests that different contexts may pick up different circumstantial accessibility relations.
(7.17) My utterance of “we could have reached Cannes” is true iff
Solange’s utterance of “we could not have reached Cannes” is false.
For if her utterance is truthvalueless as a result of indeterminacy, it is
not true, yet Solange’s is not false either. Second, we may have semantic
blindness in cases of retraction. Suppose Solange’s last utterance was
instead:

(7.18) You’re right, we could certainly have fixed it earlier. I was wrong,
we could in fact have reach Cannes.

where “I was wrong” is tantamount to saying that Solange’s previous
utterance was false. But suppose that Delphine’s second utterance has in
fact succeeded in altering the context. Solange’s first utterance may have
been true, and if it was, the retraction is mistaken. The mistake would
plausibly involve Solange’s having a false metalinguistic belief analogous
to (7.15).

Yet the view provides a reasonably good account of disagreement. On
that view, both speakers would be right in believing:

(7.19) If Delphine’s utterance is true, Solange’s utterance is not true.
It cannot be the case that both are true. So it makes sense for them not to
be satisfied with the current situation.

7.2.2 Conjunction cases

More generally, it is difficult to get two circumstantial modals with
different values in the same discourse.

Injured trombone player  Jay Jay is a trombonist who came to a party
at our house yesterday. The day after, we suddenly remember that
there is an old trombone in the attic. But Jay Jay has had a hand
injury recently.

(7.20) ?Given that we had a trombone, Jay Jay could have played
yesterday. But given his recent injury, he couldn’t have.

The utterance sounds contradictory, even though the “Given that . . . ”
clauses are meant to fix the facts respective to which each modal is to be
7.2. Circumstantial modals and real possibility

evaluated. Again, this suggests that circumstantial “could” can only take a single value at a context.

7.2.3 The proposal

We hypothesise a specific parameter of contexts:

**Real possibility** For each context $C$, the *real possibility* relation of $C$, noted $RP_C$, is a relation between worlds. When $wRP_Cw'$, we say that $w'$ is a real possibility for $w$ in $C$.

The circumstantial interpretation of modals that interests us relies on this relation. We call modals so interpreted *real possibility modals*, and we notate $COULD$ and $WOULD$ the existential one and the universal one, respectively. Their semantics is:

**Real possibility modals**

$COULDp$ is true in $C$ at $w$ iff for some $w'$ s.th. $wRP_Cw'$, $p$ is true in $C$.

$WOULDp$ is true in $C$ at $w$ iff for all $w'$ s.th. $wRP_Cw'$, $p$ is true in $C$.

(The semantics should be generalised to allow Kratzer-style ordering of the base for deontic readings, but we can ignore that here.)

$RP_C$ should be reflexive for any $C$. That is, any world is a real possibility for itself, whatever the context’s threshold for real possibility is. This validates:

**Truth is really possible** if $p$ is true in $C$ at $w$, then $COULDp$ is true in $C$ at $w$.

It would also be natural to assume that $RP_C$ is well-behaved with respect to some Lewis-style measure of closeness. That is, in every context $C$, if a world $w$ is accessible, then any world at least as close as it is accessible too: $wRP_Cw' \rightarrow \forall w'' (w'' \leq_w w' \rightarrow wRP_Cw')$ for any $w, w', C$.

---

18. “Peter could have won” typically indicates that Peter lost. One may argue that this indication of “counterfactuality” is an implication. It would follow that “Peter could have won” cannot express real possibility as we defined it. But on the implication view, one has to say that “Peter could have won” is *false* when Peter wins. That seems to me implausible. At most, “Peter could have won” presupposes that it did not win. So the “counterfactuality” indication does not prevent such claims from expressing real possibility.
7.2.4 Caveats

Two caveats are in order. First, we have called “circumstantial modals” all the interpretations of modals that a philosopher would call “alethic” (sec. 7.1.2). These include the circumstantials centred on a particular past episode that we have discussed so far. But they also include ability interpretations of modals. The first caveat is that I want to remain neutral on ability interpretations. Second, many circumstantial modals obviously involve time-relative possibilities. They may occur with multiple values within a single context. I do not want to deny this. My claim concerns only a subset of the circumstantial modals, which I call real possibility modals, that are neither ability modals nor time-relative. (Though, as I argue, real possibility modals can be considered as a limiting case of time-relative circumstantials.)

Some circumstantial modals express abilities, as in “Jay Jay can play the trombone”. On a standard and rough semantics, the sentence is true iff there is a possible world of a relevant set in which Jay Jay plays the trombone. The set of possibilities relevant for abilities may be distinct from the set of real possibilities. Here are some potential examples.

(7.21) Since Thomas lives near the lake, he could swim everyday, but he can’t swim.

In (7.21), the truth of “he could swim everyday” requires a real possibility in which Thomas swims. But the truth of “he can’t swim” requires that no ability possibility is a swimming one. If the sentence is not contradictory, the ability possibilities may not include all the real ones.

(7.22) Nixon could press the button. But all his deepest convictions would have pleaded against it. He couldn’t have done it.

In (7.22), the intended reading of “Nixon could press the button” is that Nixon had access to the button and had the physical ability to press it if he wanted. (It is not the deontic one on which he had the right to do so.) The truth of “He couldn’t have done it” requires that there is no real possibility in which Nixon presses the button. This is made impossible by his convictions. Yet the truth of “Nixon could press the
button” requires that there is an ability possibility in which he presses it. If the three sentences are not contradictory, the ability possibilities may include non-real ones.

This does not imply that the ability possibilities and the real ones are unrelated. Hacquard (2006, 104; 2010, 110) derives ability readings from the combination of a generic operator and circumstantial modality. In short, abilities are the circumstantial possibilities of worlds that are normal or ideal by some standard. Thus on its circumstantial reading, “Thomas can swim” concerns what is possible given the circumstances at our world. On its ability reading, it concerns what is possible given circumstances at normal worlds. The idea could be transposed in the real possibility framework.¹⁹

Some circumstantial modals are obviously time sensitive. For instance, (7.23) does not entail (7.24):

(7.23) Sean could have won a Oscar.

(7.24) Sean can win an Oscar.

Last year may have been the last chance for Sean to win an Oscar. If so, (7.23) is true but (7.24) is false. The latter is restricted to possibilities that share their past with actuality. So it does not quantify over all real possibilities. Moreover, the two can felicitously be combined.

It is unclear to me whether “could have” statements themselves allow time sensitivity within a single context.²⁰ In the example below, (7.25) is an uncontroversial case of time sensitivity. (7.26) applies the pattern to “could have” statements. Some find the result acceptable:

**Injured Tennis Player** Roger plays a tennis match against Rafael. The first set is a close call. But Roger is injured at the beginning of the second one, and quickly loses the match. A spectator recounts the game to a friend.

(7.25) Initially, Roger played well. But he got injured. Then he didn’t play well.

¹⁹. The transposition would probably have to rely on time-relative circumstantialials rather than on unrestricted real possibility operators. See below.

²⁰. Thanks to Jessica Leech here.
Initially, Roger could have won. But he got injured. Then he couldn’t have won.

If (7.25) is acceptable, a natural explanation is that “could have” in the third sentence quantifies only on possibilities that were open after the injury.

One issue here is a potential ambiguity in “could have”. In French, (7.26) has two distinct translations:

(7.27) Initialement, Roger pouvait gagner. Mais il s’est blessé. Alors il ne pouvait pas gagner.

Initially, Roger can-past win. But he got injured. Then he could-past not win.

(7.28) Initialement, Roger aurait pu gagner. Mais il s’est blessé. Alors il n’aurait pas pu gagner.

Initially, Roger can-past-cond win. But he got injured. Then he could-past-cond not win.

(7.27) is a past modal. (7.28) additionally involves the conditionnel morphology, which typically signals counterfactuality. Now I find (7.27) roughly acceptable, but (7.28) sounds to me like a retraction. This suggests that conditionnel possibilities are not time sensitive, while past modal possibilities are.

An investigation of these issues is beyond the scope of this work. Here I will simply sketch a picture that I will work with.

A time-sensitive circumstantial like (7.24) quantifies over worlds whose history matches actuality up to a given time — in that case, the present. But it does not quantify over all such worlds that a typical philosopher considers metaphysically possible. Otherwise hardly any ordinary possibility claim would be false. Some further restriction is needed. The

21. In French, conditionnel marks subjunctive conditionals as opposed to indicative ones.


23. Condoravdi (2001, 13) calls the possibility involved in circumstantials “metaphysical”. In philosophical contexts, the phrase would be misleading. Hence our term “real possibility”.
suggestion I make is that a time-sensitive modal results from the intersection of a certain history with a background set of possibilities, namely the real possibilities. Write $\text{COULD}(t)p$ for the time-sensitive circumstantial. Write $H(w, t)$ for the history of $w$ up to $t$. The proposal is:

**Time-sensitive circumstantial** $\text{COULD}(t)p$ is true in $C$ at $w$ iff there is a $w'$ such that $wRP_Cw'$ and $H(w, t)$ is true in $w'$, and $p$ is true in $C$ at $w'$.

Following the suggestion above, past and present modals would express the time-sensitive circumstantial, while *conditionnel* modals would express the unrestricted real possibility circumstantial. The English “could have” would be ambiguous between the two. 24

The real story may be more complicated than that. 25 What matters for our present purposes is that (a) some circumstantial modals express unrestricted real possibility, (b) other circumstantial modals indirectly rely on the real possibility set. Our criteria for real possibility modals is that they resist multiplicity of values in a single context and that they are typically expressed with the *conditionnel* form in French.

### 7.3 Subjunctive conditionals and real possibility

In this section, we give the second argument for the real possibility parameter: that it is independently needed in a plausible semantics for counterfactuals.

24. From the semantics it follows that $\text{COULD}(t)p \rightarrow \text{COULD}p$ is valid. Some rough tests:

- “He could win. Therefore he could have won.”
- “Yesterday he could buy a car. So he could have bought a car.”
- “He can win. Therefore he could have won.”

The latter is infelicitous because of the counterfactuality implicature of “he could have won”, but not clearly invalid.

25. For instance, the imperfective aspect of past tense in French is versatile, and one of its values seems equivalent to the *conditionnel* (Hacquard, 2006, 82–6).
7.3.1 von Fintel and Gillies on subjunctive conditionals

On the standard Lewis-Stalnaker view, subjunctive conditionals are variably strict conditionals (Stalnaker, 1968; Lewis, 1973, §1.3, see appendix A for details). The truth of a subjunctive conditional $p \Box \rightarrow q$ requires the truth of the material conditional $p \rightarrow q$ over a domain of worlds — hence, strictness. But the domain of worlds over which the truth of the material conditional varies with the antecedent proposition — hence, variability. Their proposal is that $p \Box \rightarrow q$ holds at $w$ iff the $p$-world closest to $p$ are $q$ worlds.

The main motivation for the Stalnaker-Lewis semantics comes from felicitous sequences of conditionals — called Sobel sequences after Sobel (1970) — such as:

(7.29) (a) If Sophie had gone to the parade, she would have seen Pedro dance; but of course,

(b) if Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance. (Gillies, 2007, 331)

If the two conditionals operated on the same set of possibilities, they would be contradictory. By the first, there is no world in the set in which Sophie goes to the parade and fails to see Pedro dance. By the second, there is such a world. But the two counterfactuals do not appear to be contradictory. On a variably strict semantics, the two conditionals may operate on different sets of worlds, since they have different antecedents. Thus they may both come out true.26

Irene Heim has pointed out, however, that reverse Sobel sequences are infelicitous (reported by von Fintel, 2001, 130):

(7.30) (a) If Sophie had gone to the parade and been stuck behind someone tall, she would not have seen Pedro dance; but of course,

26. In other terms, Strengthening the Antecedent fails: $p \Box \rightarrow q$ does not entail $p \& r \Box \rightarrow q$ for all $r$. 

7.3. Subjunctive conditionals and real possibility

(b) if Sophie had gone to the parade, she would have seen Pedro dance. (Gillies, 2007, 332)

The sequence sounds contradictory. However, on the variably strict semantics, both conditionals in the reverse sequence are true if and only both are true in the original one. The infelicity is puzzling.

Von Fintel (2001) and Gillies (2007) avoid the problem by adopting a context-sensitively strict semantics (see appendix A for details; the view was considered by Lewis, 1973, 13 and Stalnaker, 1984, 123–4). On their view, the domain of possibilities over which subjunctive conditional operate is provided by context, and does not vary with their antecedent. von Fintel (2001, 127) calls that parameter the modal horizon of the context, Gillies (2007, 330) calls it the counterfactual score of the context. The view readily explains the infelicity of reverse Sobel sequences. If (7.30 a) is true, there is a world in the domain in which she goes to the parade and fails to see Pedro dance. But (7.30 b) is true only if there is not such a world. So the sequence is contradictory.

Now the semantics has a problem with the original sequence, since if the domain is constant, the sequence is contradictory. But von Fintel and Gillies also argue that subjunctive conditionals affect the context, as follows:

**Entertainability presuppositions** An assertion of $\Box p \rightarrow q$ presupposes that there is a $p$-world in the counterfactual domain. (Gillies, 2007, 333)

The assumption is natural: the same phenomenon is observed with universal quantifiers. “All the children are asleep” presupposes that there are children in the contextually relevant domain. This provides an account of the original sequence. When (7.29 a) is uttered, there is no world in the counterfactual domain of the context in which she is stuck behind someone tall at the parade. When (7.29 b) is uttered, this forces some such worlds to be included in the domain. We get into a new context, in which

---

27. The differences between von Fintel’s and Gillies’s views concern the details of semantic dynamics and *might* conditionals (Gillies, 2007, 342–8). We ignore them here.
(7.29b) is true. In the new context, (7.29a) is false, and thus it cannot be uttered again felicitously. The expansion is not made arbitrarily. The p-worlds introduced are the closest ones on some Lewis-Stalnaker-style ordering. Including them also forces the inclusion of all worlds that are as close as them by the ordering.

7.3.2 Unified real possibility semantics

The semantics is attractive and provides a plausible solution to the puzzle. I will adopt it without further discussion here. It posits a contextual parameter, the counterfactual domain, over which subjunctive conditionals quantify. We propose to identify the counterfactual domain and the domain of real possibilities of the world of utterance. Real possibility modals and subjunctive conditionals operate over the same set of possibilities at every context.

The identification is not entirely straightforward. Real possibility is an accessibility relation. von Fintel’s and Gillies’s counterfactual domain is a set of possibilities. We identify von Fintel’s and Gillies’s counterfactual domain with the set of real possibilities of the world of evaluation. The difference shows up in unusual and hard-to-evaluate contexts — counterfactuals evaluated at other worlds than the world of utterance. I think that von Fintel’s and Gillies’s sets are likely to run into problems analogous to the one that leads Lewis’s account of “know” to violate factivity. But I will not discuss the point further here.

The resulting picture is the following. We write COULD for the real possibility modal, and WOULD for the real necessity modal. We formally represent them as generalised quantifiers.

---

29. The idea is suggested in von Fintel (2001, 140) and Gillies (2007, 355) for “would” modals, e.g. “That would never happen.” If English would modals are cousins of the French conditionnel, and if the latter expresses real possibility, as suggested above, that is expected.
30. Try: “It could have been the case that if it rained, we would have been wet, but fortunately it was not.”
31. \((\text{SOME} : Fx)(Gx) := \exists x(Fx \land Gx)\) and \((\text{ALL} : Fx)(Gx) := \forall x(Fx \rightarrow Gx)\).
7.3. Subjunctive conditionals and real possibility

Real possibility modals

\( \text{COULD}p \) is true in \( C \) at \( w \) iff for some \( w' \) such that \( wRP_Cw' \), \( p \) is true in \( w' \) in \( C \).

\( \text{WOULD}p \) is true in \( C \) at \( w \) iff for all \( w' \) such that \( wRP_Cw' \), \( p \) is true in \( w' \) in \( C \).

\[ \llbracket \text{COULD}p \rrbracket^C = \{ w : (\text{SOME} : wRP_Cw')(w' \in \llbracket p \rrbracket^C) \} \]

\[ \llbracket \text{WOULD}p \rrbracket^C = \{ w : (\text{ALL} : wRP_Cw')(w' \in \llbracket p \rrbracket^C) \} \]

We take the subjunctive conditional to be a restricted version of the \( \text{WOULD} \) generalised quantifier.\(^{32}\) We write \( (\text{WOULD} : p)(q) \) for “If \( p \), would \( q \)”. We have:

Real possibility conditional

\( (\text{WOULD} : p)(q) \) is true in \( C \) at \( w \) iff for all \( w' \) such that \( wRP_Cw' \) and in which \( p \) is true, \( q \) is true in \( w' \) in \( C \).

\[ \llbracket (\text{WOULD} : p)(q) \rrbracket^C = \{ w : \text{ALL}(wRP_Cw' \land w' \in \llbracket p \rrbracket^C)(w' \in \llbracket q \rrbracket^C) \} \]

Bare modals are a limiting case of conditionals.

7.3.3 Real and theoretical possibilities

Let us look again to the disagreement case (p. 7.12) in light of the unified view. Recall:

(7.31) (a) Delphine. What a shame. We could have reached Cannes!

(b) Solange. No way! With the motor in that shape, we couldn’t have reached it.

(c) Delphine. But the motor could have been in a better shape. If it had been, we would have reached Cannes. So we could have reached Cannes!

(d) Solange. Well if the motor had been in a better shape, we could have reached Cannes. But it wasn’t. So we couldn’t have reached it.

At (7.31 c), Delphine is trying, so to speak, to expand the real possibility set to worlds in which they reached Cannes. She does so by asserting a bare

\(^{32}\) See Kratzer (2010a) for a defence of the generalised quantifier view of conditionals.
modal, which would introduce worlds in which the motor is in better shape, and a subjunctive conditional, which states that those worlds are ones in which they reach Cannes. Delphine could have done the same by simply asserting the conditional, relying on the Entertainability presupposition of the conditional:

(7.32) If the motor have been in a better shape, we could have reached Cannes. So we could have reached Cannes!

This suggests two rules for the dynamics of real possibility:

**Accommodating the possibility modal** If COULDp is asserted in C, C tends to be such that RP_C counts some p-worlds as real possibilities of the world of evaluation.

**Entertainability presupposition of the conditional** If (WOULD : p)(q) is asserted in C, C tends be such that RP_C counts some p-worlds as real possibilities of the world of evaluation.

At (7.31 d) Solange tries, so to speak, to reduce the domain of real possibilities. By analogy with accommodation of the possibility modals, we could consider the rule:

**Accommodating the necessity modal** If WOULDp is asserted in C, C tends to be such that RP_C excludes not-p-worlds from the real possibilities of the world of evaluation.

The rule has some plausibility. Suppose Solange had said instead:

(7.33) The motor couldn’t have been in a better shape, it is way too old!

(7.34) Even if we had tried to fix it, the motor would still have broken down at some point.

Both utterances appear to restrict the set of real possibilities.33

Interestingly, the set of real possibilities also appears to be affected by factual claims. Thus in (7.31 d), what Solange in fact does is to mention that a possibility on which their reaching Cannes depended — namely, the car being in good shape — does not actually obtains. Something like the following appears to be at play:

---

33. See also von Fintel (2001, 140) and Gillies (2007, 355).
Possibility dependence  Where $q$ depends on the obtaining of $p$, asserting not-$p$ in $C$ tends to restrict $RP_C$ so that $q$ worlds are excluded from the real possibilities of the world of valuation.

Together the rules create unstable modal contexts. Suppose that $p$ is false, and that $q$ depended on $p$. Two opposite lines of reasoning are available:


Only if $p$, would $q$. Not-$p$. So not-$COULDq$.

Delphine adopts the first, and Solange the second. My guess is that in these sorts of situation the debate will naturally focus on the possibility of the conditions on which the disputed possibility depends. Thus Solange will soon be pressed to argue that the car couldn’t have been in a better shape. Delphine will be forced to justify her claim that it could. The debate may then turn to the preconditions of the car’s being in bad shape, and so on.

* Mentioning* possibilities that are out of the domain of real possibilities does not suffice to turn them into real ones. (7.31 d) offers a striking example: Solange’s reasserts the conditional, which normally presupposes that its antecedent is true in some real possibility. But she treats that possibility as a merely theoretical one, as the latter part of her assertion makes clear. More generally, we call the the non-real possibilities of a context the (merely) theoretical possibilities. We may say that a conditional is theoretical if its antecedent is out of the set of real possibilities. By default, our semantics makes all theoretical conditionals trivially true — in the fashion of “when pigs fly” conditionals. A subtler treatment may be required, but I will not go into that question here.

7.4 The real possibility account of knowledge attributions

In this section, we develop a version of epistemic contextualism based on the method infallibilism idea, paired with the idea that knowledge is
a matter of avoiding real possibilities of error. The resulting view avoids the problems faced by DeRose’s and Lewis’s views. It does not posit a context-sensitive mechanism specific to “know”. In particular, it does not rely on a Rule of Sensitivity, though it partly vindicates sensitivity conditionals. It provides a more satisfying account of what taking a possibility seriously consist in. It accounts for the efficacy of both first-person and third-person arguments and avoids violation of closure. In short, it is the account any modal epistemic contextualist should adopt.  

7.4.1 Semantics of “Know”

Method infallibilism is the claim that:

Method infallibilism  $S$ knows that $p$ iff $S$ believes that $p$ on the basis of a method that could only yield true beliefs.  

Method infallibilism is turned into a contextualist semantics if we take the “could” to express a context-relative notion of real possibility. Both sides of the equivalence turn out to be context sensitive. Assuming that the equivalence holds at every context, we derive a contextualist semantics. We formulate it as an amendment of (see 6.4.3):

Real possibility account of “know”  For all $C$, $\alpha$, $K_C \alpha$ iff $K_C \alpha$ iff for some method $m$: $Bma$ and for all $\beta$ such that $w_\alpha \text{RP}_C w_\beta$, $\neg E_m \beta$.

(where $E_m \beta = Bm\beta \land \neg T\beta$, and $RP_C$ expresses real possibility in $C$.)

That is: relative to a context $C$, “know” refers to a property that is satisfied by a case $\alpha$ iff there is a method $m$ such that $\alpha$ is a case of belief based on $m$ and $m$ is infallible over the real possibilities of $w_\alpha$ in $C$.

34. Ichikawa (ms.a) defends essentially the same view. He adopts a contextualist semantics of counterfactuals and a modal epistemic contextualist view of “knowledge” attributions, and claims that the same set of possibilities governs both. However, he relies on an amended version of Lewis’s account that faces the necessary truths problem (6.4), and his semantics is not developed enough to tell whether he faces the violation of factivity problem (6.6.2). Moreover, he uses the safety and sensitivity conditionals as an argument for the unified view. We rather point out that the unified view, paired with a method-based account, predict the failures of the sensitivity and safety conditionals. Finally he does not connect the semantics with bare modals.

35. Cf. 4.2.5.
7.4.2 Taking a possibility seriously

Additionally, we propose that:

**Real possibility account of taking seriously** Taking a possibility seriously consists in *taking it to be a real possibility*.

The proposal requires comment. So far, I have identified real possibilities as a contextual parameter in the semantics of modals and subjunctive conditionals. But by saying that taking a possibility seriously consists in taking it to be a real possibility, I do not mean that taking a possibility seriously consists in holding some metalinguistic belief about the semantics of modals. Rather, we have a group of inter-defined notions here. Taking a possibility \( p \) seriously consists in believing that \( p \) could have happened. A belief that \( p \) could have happened is true iff in the belief’s context, \( p \) is a possibility that one ought to take seriously. One ought to take a possibility \( p \) seriously in a context, if it is true in that context that \( p \) could have happened.\(^{36} \), \(^{37} \)

---

\(^{36}\) The claims are obviously false on some readings of “ought”. If I mistakenly believe that \( p \), there is perhaps a sense in which I ought to take \( p \) seriously, even though \( p \) might not be a real possibility in my context.

\(^{37}\) Revision note. Timothy Williamson pointed out serious difficulties for the account. First, one can take a possibility seriously without believing that it is a real possibility. For instance, one can take seriously the possibility that the world is Newtonian (obeys to Newtonian laws), without believing that it is physically possible that the world is Newtonian. Rather, one believes that if the world is Newtonian, it necessarily is, and if it is not, it necessarily is not. So one at most believes that it is (epistemically?) possible that it is a real possibility. More strikingly, one can take seriously both the possibility that Hesperus is identical to Phosphorus and that the possibility that they are distinct. But one cannot both believe that it is a real possibility that they are distinct and a real possibility that they are identical: for assuming the necessity of identity, if there are identical, there is no real possibility that they are distinct, and if they are distinct, there is no real possibility that they are identical. Second, it may be false in a context that a possibility is real, and yet one ought to take it seriously. For instance, if Hesperus is not Phosphorus, it is false that the possibility that Hesperus is not Phosphorus is real, and yet there may be contexts in which one ought to take seriously the possibility that they are distinct.

As Williamson also notes, this has consequences for my account of the force of sceptical arguments below. The account is that: (a) if one takes a possibility seriously, one believes that it could have happened (it is a real possibility), and (b) if the sceptical scenario is a real possibility of error on the basis of the same methods, then I do not know. So taking a sceptical scenario seriously commits one to believing that one does
Moreover, one should be careful about the sloppy term of “possibility” (see sec. 6.5). *Real possibilities* in my sense are primarily possible worlds. When we say, relative to a context and a world, that a proposition $p$ is a real possibility, we mean that some really possible worlds of that context and world are $p$-worlds. But when we *take a possibility seriously*, the possibility in question is a proposition. So strictly speaking, where $p$ is a proposition, the proposal is:

Taking a possibility $p$ seriously consists in believing that some really possible worlds are $p$-worlds.

The paradigmatic way of believing that some really possible worlds (of one’s world and context) are $p$-worlds is to entertain a belief of the form “*COULD*p”.

Finally, note that we do *not* endorse the following:

there is a real possibility that $p$ in a context $C$ iff $p$ is taken seriously in $C$.

To take a possibility seriously is to *think* that a possibility is a real one. It is not to make it a real one. While the set of real possibilities of one’s world in one’s context may depend in part in one’s interests, beliefs, what one attends to, and so on, it is not assumed that the set is whatever the speaker thinks it is. Real possibilities typically are fairly opaque to agents.

A relevant point to note here is that real possibilities are not directly given as a set of possibilities not know. But if as Williamson argues, (a) fails, the account fails.

At this stage I can only gesture towards a solution. Suppose that one takes seriously the possibility that ($p$) Hesperus is not Phosphorus, in the sense relevant to knowledge attributions. (If that is a third-person attribution, one may very well believe that Hesperus *is* Phosphorus.) As Williamson argues, this cannot mean that one believes that $\neg p$ is a real possibility. But, I suggest, this means that one believes that $\neg p^*$ is a real possibility, where $p^*$ is relevantly similar to $p$. The relevant similarity is a matter of the subject’s methods, not of the propositions themselves: in a case of successful sceptical hypothesis (again, be it first- or third-person), $p^*$ will be a possibility such that methods that would lead the subject to believe that $p$ would lead her to an error about $p^*$. Thus when one “takes seriously” the possibility that Hesperus is not Phosphorus, one believes that there is a real possibility involving two distinct planets in which the method available to the subject on the basis of which she believes or would believe that Hesperus is Phosphorus would mistakenly lead her to believe that these planets are identical in that real possibility.
7.4.3 Comparison with Lewis’s and DeRose’s accounts

The Real Possibility account has a number of advantages over Lewis’s and DeRose’s views.

Factivity

First, it validates factivity, contra Lewis (6.6.2). Every world is a real possibility for itself (sec. 7.2.3). Knowledge at a world requires avoidance of error at the real possibilities for that world. So knowledge is incompatible with false belief at that world.

Non-automaticity

Second, it does take any mention of a sceptical scenario to produce sceptical contexts, contra Lewis (6.5). Consider Blome-Tillmann’s example:

Imagine you saw your teenage son sneaking away through the window late at night. When you confront him the next morning he replies somewhat desperately: ‘How do you know I left the house? I mean, for all you know you might have dreamt it. It was late at night, wasn’t it? (Blome-Tillmann, 2009b, 246–7)

On the present view, the son’s attempt at a sceptical argument is unsuccessful insofar it fails to establish that there was a real possibility of the parent’s dreaming the escape. And it is unsuccessful at raising doubt insofar as it fails to make us think that there was such a real possibility.

No Rule of Sensitivity

Third, it avoids problems that beset DeRose’s rule of sensitivity. The first problem is Blome-Tillmann’s (2009a, 387) example (sec. 6.4.3):

(7.35) I know that nothing can travel faster than light.

Applying the Rule of Sensitivity to (7.35) brings into consideration worlds that are wildly different from ours. The utterance should thus be of a kind that generates sceptical intuitions, but it is not. On the Real Possibility
approach, we drop the Rule of Sensitivity. (7.35) does not impose real possibilities in which things travel faster than light. This does not make the knowledge claim (7.35) trivially true either, because the subject may believe that nothing can travel faster than light on a fallible basis.

The second problem with the Rule of Sensitivity is its puzzling failure to raise standards when considering victims of Evil Demons (sec. 6.6.4). When we consider a subject’s belief that \( p \), says the Rule, we have to raise standards for “knowledge” high enough so that the belief is knowledge only if sensitive. Now if we are considering a sceptical scenario in which René has a false belief because he is victim of an Evil Demon, however low the standards are, they are low enough to require sensitivity of him. So consideration of the scenario fail to raise the standards. But the sceptical force of such scenarios was precisely what DeRose (1995) set out to explain. By contrast, on the real possibility account, the explanation is straightforward. When presented with the scenario, one is led to believe that there is a real possibility of someone — not necessarily oneself — being such a victim. Provided that the victim’s beliefs are based on the same methods as ours, this implies that our methods are fallible.\(^{38}\) Thus we conclude we do not know.

A third problem for DeRose’s Rule of Sensitivity was to find analogues in natural language (sec. 6.3.5). The best analogy we found was the following: upon considering René’s predicament, we wonder whether it may be ours. The problem with the analogy was that it only predicted that we would react to sceptical arguments by saying “Well, perhaps I do not know, then”, and not “I do not know, then”. On the present view, sceptical scenarios act simply as counterexamples to a universal claim. If one knows that \( p \) on some basis, then error on that basis is avoided at all really possible cases. But the sceptical scenario, if taken seriously, shows that there is a possible case in which error is not avoided. The prediction is that the proper reaction is “I don’t know” and not “Perhaps I do not know, then”.

\(^{38}\) As we pointed out sec. 4.3.5, the argument assumes internalism.
Third person sceptical arguments

Fourth, the new account of taking possibilities seriously avoids the problem of third-person sceptical arguments (6.5.3). Sceptical scenarios are as efficient at generating knowledge denials in the third person as in the first. The conclusion is more unsettling in the latter case, but the unsettling aspect of first-person scepticism is not essential to the argument. The belief account of *taking seriously* fails to predict the efficacy of third-person sceptical arguments. By contrast, on the present view, real possibilities of error undermine knowledge ascriptions even when the ascriber knows that they do not obtain. Take the third-person painted mule case, for instance (6.5.3). Bouvard takes himself to know that the animal in the pen is a genuine zebra. Pécuchet leads him to consider the possibility that it is a painted mule as a real possibility — without leading him to doubt that it is *non*-actualised. Still, this is sufficient for him to think that there is a real possibility in which the visitor’s methods go wrong, and thus to deny that the visitor knows. Recognising the painted mule possibility as a real one may ultimately undermine Bouvard’s own claim to know, but it need not.

No linguistic exception

Finally, the account avoids positing a context-sensitive mechanism that is specific to “know”. Real possibilities are independently needed for modals and subjunctive conditionals. This allows the account to reply to an objection to contextualism raised by Stanley (2005b, 57–68). Stanley points out that context-sensitive expressions normally allow multiple value within the same sentence. Thus in some contexts (7.36) can express the proposition that every sailor of one ship waived to every sailor of another ship:

(7.36) Every sailor waved to every sailor. (Stanley and Williamson, 1995, 294)

Stanley (2005b, 57) suggests the generalisation that all context-sensitive expressions allow such shifts. Epistemic contextualists do not want to
allow such shifts. They undermine the contextualist’s explanation of closure-based sceptical arguments (see below sec. 7.4.4). So they are put in the uncomfortable position of having to argue that “know” is a linguistic exception.

The real possibility account avoids the problem by pointing out two linguistic constructions that resist such shifts: real possibility modals and subjunctive conditionals. Stanley (2005b, 53) argues that modals allow shifts within a single modal flavour, but I find his argument unconvincing. He puts forward the following dialogue:

(7.37) A. It’s possible to fly from London to New York City in 30 minutes.
   B. (overhearing) That’s absurd! No flights available to the public today would allow you to do that. It’s not possible to fly from London to New York City in 30 minutes.
   A. I didn’t say it was. I wasn’t talking about what’s possible given what is available to the public, but rather what is possible given all existing technology.

Now Stanley’s dialogue is formulated with the complex modal “it is possible that”. It is hasty to generalise from the complex modal to the simple one can: the complex is made out of a gradable adjective, and the adjective may provide a context-sensitive parameter that the simple modal cannot. If we replace “it is possible that” by “can”, the dialogue does not sound felicitous to me. Moreover, even if it was felicitous, one should check that the alleged context-sensitivity does not boil down to a difference in complement clause. Thus (7.38) is better analysed as (7.39) than as a context shift of can:

(7.38) With an A310 plane, one cannot fly from London to New York.
   But with an A340, one can.

(7.39) One cannot (fly from London to New York with an A310 plane).
   But one can (fly from London to New York with an A340 plane).

Thus Stanley’s example is not conclusive. By contrast, we have pointed out cases which suggest that could have modals resist mid-sentence shifts
(sec. 7.2). The real possibility view avoids special pleading for know.

### 7.4.4 Knowledge ascriptions and subjunctive conditionals

The real possibility account predicts some systematic links between knowledge ascriptions and subjunctive conditionals. Yet the account avoids problems that Nozick’s and Sosa’s subjunctive conditional requirements on knowledge faced.

The real possibility account implies that in any context in which (7.40) is true, (7.41) is true for some m and any q:

(7.40) $S$ knows that $p$.

(7.41) $S$ would not believe $q$ on the basis of $m$ while $q$ is false.

$\text{WOULD } (\neg (S \text{ believes that } q \text{ on the basis of } m \text{ and } q \text{ is false}))$.

Given the unified semantics (sec. 7.3.2), (7.41) entails (7.42), for any context, $m$, $q$:

(7.42) If $q$ was false, $S$ would not believe $q$ on the basis of $m$.

$(\text{WOULD} : \neg q)(S \text{ believes that } q \text{ on the basis of } m)$

In particular, in any context in which (7.40) is true, there is an $m$ such that (7.43) is true:

(7.43) If $q$ was false, $S$ would not believe $q$ on the basis of $m$.

Which is Nozick’s method-relative sensitivity conditional.$^{39}$

Moreover, subjunctive conditionals contrapose on the context-sensitively strict semantics we adopt (see app. A). Given the unified semantics, (7.41) entails (7.44), for any context, $m$, $q$:

(7.44) If $S$ believed $q$ on the basis of $m$, $q$ would be true.

$(\text{WOULD} : S \text{ believes } q \text{ on the basis of } m)(q)$.

In particular, in any context in which (7.40) is true, there is an $m$ such that (7.45) is true:

---

$^{39}$ More precisely, Nozick’s method-relative sensitivity conditional revised as Luper-Foy and Williamson have suggested. See sec. 3.5.3 for details.
If $S$ believed that $q$ on the basis of $m$, $q$ would be true.

Which is a method-relative version of Sosa’s Safety conditional (sec. 3.4.5).

This is good news, because the Sensitivity and Safety conditionals provide plausible diagnosis of why knowledge obtains or fails in many cases, especially Gettier cases. Nevertheless, both the Safety and the Sensitivity conditionals face problems. However, the problems are avoided in the real possibility account.

On their simpler versions — without reference to methods —, the Safety and Sensitivity conditionals face easy counterexamples such as Nozick’s grandmother (Nozick, 1981, 179). To give another example: Jack is pathologically convinced that his neighbour James is constantly watching him. Usually he is wrong, but right now Jack sees that James is watching him. Jack knows that James is watching him, even though (1) if James was not, Jack would still believe that he was, (2) it is not the case that if Jack believed that James watched him, James would watch him. Both the simple sensitivity conditional and simple safety one fails. But their method-relative versions avoid the problem.

Sensitivity, but also Safety, lead to widespread violations of deductive closure if we assume a variably strict semantics for conditionals (see sec. 3.4.5 and app. A). But in the present account we do not, so the failures are avoided. In particular, the account can endorse the standard contextualist diagnosis of closure-based formulations of sceptical arguments (see e.g. DeRose, 1995). Consider:

$(7.46)$ $S$ does not know that the animal is not a painted mule.

$(7.47)$ If $S$ does not know that the animal is not a painted mule, $S$ does not know that the animal is a zebra.

$(7.48)$ $S$ knows that the animal is a zebra.

In no context $(7.46)$–$(7.48)$ are all true. In those in which it is a real possibility that the animal is a painted mule, $(7.48)$ holds and $(7.46)$ fails.

---

40. Goldman’s Dachshund case and Williamson’s slight insensitivity problem (cf. Williamson, 2000, 159–60 on both) are also handled by the method qualification, provided we use our intuitions about knowledge to individuate methods, as in sec. 4.3.4.
In a context in which it is not a real possibility, (7.48) may hold and if it does, (7.48) fails.

Nozick’s sensitivity conditional is trivially satisfied by beliefs in necessary truths (Nozick, 1981, 186). The method infallibilist account avoids this by implying a range of conditionals. For instance, consider our Prime Number case (sec. 4.1.1). Given that Primo would use the same methods on 47 and 49, for him to know that 47 is prime, the two following conditionals must hold:

(7.49) If 47 was not prime, Primo would not believe that it was.
(7.50) If 49 was not prime, Primo would not believe that it was.

The second fails, because there is a real possibility in which 49 is not prime and Primo believes it is.41

Nozick’s account implausibly denies most inductive knowledge. Consider Vogel’s Ice Cubes case (Vogel, 1987, 206): I have left ice cubes in glass out in the direct sun several hours ago, and have not seen them since. I know that they have melted. Yet the following holds, for any method $m$ on which this belief is based:

(7.51) If the ice cubes had not melted, I would still believe that they had on the basis of $m$.

For (7.51) it apparently follows that:

(7.52) It is not the case that if the ice cubes had not melted, I would not believe that they had on the basis of $m$.

That is, Nozick’s method-relative sensitivity conditional is false. But the present account also endorses Nozick’s method-relative sensitivity conditional. So if (7.52) is true in the current context, it is false to say that I know that they have melted.

On the real possibility view, the problem is avoided by saying that the possibility that the ice cube have melted is not a real possibility in the current context. This means that (7.51) is a theoretical conditional (sec. 7.3.3).

41. Note that on a strongly centred semantics (the actual world is the unique world maximally close to itself), (7.50) holds as long as Primo does not actually believe that 49 is prime.
In the unified semantics of sec. 7.3.2, this implies that (7.52) does not follow from (7.51). In fact, in a context in which there is no real possibility that the cubes have not melted, both (7.51) and (7.53) hold:

(7.53) If the ice cubes had not melted, I would not believe that they had on the basis of m.

These are “when pigs fly” conditionals in the context.

Alternatively, one may endorse a subtler semantics for theoretical conditionals, on which (7.51) and (7.53) fails. But then, given our semantics for “know”, theoretical conditionals such as (7.51) will not be entailed by the claim that I know that the ice cube have melted. Either way, the account avoids the induction problem.

7.4.5 Conclusion

We have formulated a contextualist semantics for “know”. The semantics is based on the method infallibilist account of knowledge. To cash out infallibility, it relies on the notion of real possibility at play in the semantics of circumstantial modals and subjunctive conditionals. The semantics says that “knowledge” requires a method that avoids real possibilities of error. What counts as a real possibility is assumed to be a context-sensitive matter. We have argued that the view avoids a number of problems that affect Lewis’s and DeRose’s contextualist accounts. The semantics also predicts systematic links between knowledge ascriptions and subjunctive conditionals. But it does so without running into the problems that the analyses of accounts in terms of subjunctive conditionals have faced.

7.5 Contextualism and invariantism

The contextualist picture we have arrived at is attractive. In this section I consider whether we should endorse it. I do not think that the question is easily settled. There are a number of pressures against contextualism about knowledge. I will briefly discuss three. The first
is that if the contextualist story is right, the property “know” refers to in ordinary discourse is not interesting to epistemologists. The point expands on Sosa (2000b, 2). The second is that the contextualist story implies that we are semantically blind in implausible ways (Schiffer, 1996; Hawthorne, 2004, 98–111; Stanley, 2005b, 57–68; Williamson, 2005a). The third is that contextualism severs the link between knowledge and action (Hawthorne, 2004, 86, Williamson, 2005b, §4).

My conclusion is conditional. Given the intuitive links between knowledge ascriptions, subjunctive conditionals, and modals, the unified semantics is desirable. However, the unified semantics is not essentially contextualist. It is open to consider that real possibility is constant across contexts. My suggestion is thus that contextualism about knowledge and about real possibility stand or fall together. Invariantism is an option, but it requires us to go against widely held views on modals and subjunctive conditionals.

7.5.1 Naturalness

Sosa (2000b, 2–4) argues that contextualists about knowledge commit a contextualist fallacy: they fallaciously infer claims about knowledge from claims about the correct use of “know”. This is prima facie surprising, since it seems that one knows iff it is correct to say that one “knows”, that is:

\[ S \text{ knows that } p \text{ at } t \iff \text{ it is correct for anyone to say “} S \text{ knows that } p \text{ at } t \”. \]

42

Barring exceptions such as non-literal uses, regimented uses, and so on, this sounds right. But precisely, if “know” is indeed context-sensitive, the principle fails. So it is open to grant that some use of “know” is true, without accepting that there is knowledge. Thus the semantics of knowledge ascription is at least not directly relevant to epistemology.

To this contextualists would reply that this only shows that knowledge claims made in our current context — the context of epistemology —
may be true while the corresponding claims made in ordinary contexts are false. This raises two points. First, one may think that the object of epistemology is precisely what is ordinarily called “knowledge”. Second, if it is not, one may wonder why the epistemologist’s reference of “know” should be more important than the ordinary one.

This, I think, raises an important issue. Contextualist theories of knowledge tend to ascribe an unnatural referent to “know” in ordinary uses. For instance, if we amend Lewis’s semantics to avoid the factivity problem, we may get a semantics according to which “know” requires the avoidance of error at worlds that are either (a) close to the subject or (b) attended to by the speaker. If that is indeed the semantics of ordinary “knowledge” ascriptions, then one wonders why it should be worth studying, except to learn about our concepts — that is, as a part of psychology. Lewis suggests as much:

What is it all for? Why have a notion of knowledge that works in the way I described? (Not a compulsory question. Enough to observe that we do have it.) But I venture the guess that it is one of the messy short-cuts - like satisficing, like having indeterminate degrees of belief - that we resort to because we are not smart enough to live up to really high, perfectly Bayesian, standards of rationality. (Lewis, 1996, 563)

In those conditions, it would be very tempting to devise a notion — call it neutrally knowledge* — that is natural but close enough to the original one. For instance, knowledge* may require avoidance of error at close worlds only. The notion is likely prove much more interesting to the epistemologist than the apparently arbitrary disjunction. It may afford more robust generalisations, and have significant properties.

In this sort of situation we can regard the theoretical language as a regimented language, slightly distinct from ordinary language. “Contextualism about knowledge” becomes ambiguous. The phrase may mean that “know” in ordinary language is context-sensitive. But it may also mean that the regimented language of theory is or should be context-
sensitive. The latter kind of view makes sense if instead of finding one natural *knowledge* property, one finds a range of them, for instance. We can distinguish the two views by the labels *descriptive contextualism* and *normative contextualism*.

On the kind of arbitrary-looking semantics that Lewis endorses, contextualism can at most be descriptive. But the real possibility account can be put forward as a normative contextualist account. Real possibility is presumably natural enough, the account is simple, and has a well defined structure (ch. 5). It provides us with a range of natural knowledge-properties, and we should allow our theoretical discourse to contextually range over them.

So the unnaturalness and irrelevance worries for contextualism do not affect the real possibility account.

### 7.5.2 Semantic blindness

A speaker is semantically blind if she has mistaken metalinguistic beliefs as a result of her failing to notice context shifts. As we have seen, contextualists about circumstantial modals have to endorse a certain amount of blindness (sec. 7.2). Contextualists about “*know*” have to as well (Schiffer, 1996; Hawthorne, 2004, 98–111; Stanley, 2005b, 57–68; Williamson, 2005a). The alleged semantic blindness shows up in *disagreement* and *retraction* cases in particular, as we illustrated with circumstantial modals (sec. 7.2).

False metalinguistic beliefs bring false first-order beliefs. One year ago, I said that I “*knew*” that *p*. I remember that what I said was true. I also believe that *if it was true to say that I “*knew*” that *p* then, I *did know* that *p* then*. However, that belief is false, because my current context is different from the one in which I made the original assertion. I infer that I knew that *p*. But that is false. Generalising, contextualism with semantic blindness endangers the transmission of information across contexts (Williamson, 2005b, §4).43

---

43. *Revision note.* François Récanati objected that we may have devices of “hybrid
DeRose’s (1992, 927) defence of semantic blindness is that our intuitive first-order judgements involving “know” are more reliable than our metalinguistic judgements about whether an utterance involving “know” is true. But as Williamson (2005a, 223) argues, metalinguistic judgements may be as important in fixing the semantic value of a term as the first-order ones. For metalinguistic judgements screen out misapplications of a term. The fact that speakers are inclined to think that they were wrong in calling something a “dog” when they later discover that it was not a dog contributes to the fact that “dog” refers to dogs and not to dog-looking objects.

Our tendency to disquote, retract and disagree when using “know” suggests that we simply do not take “know” to be a context-sensitive term. It may be that we apply “know” in a context-insensitive manner, whenever evaluating “knowledge” claims across contexts, we treat them as we would if we believed them not to be context-sensitive. Let us summarise these aspects of our behaviour by saying that we linguistically intend “know” to be context-insensitive.

Linguistically intending an expression to be context-insensitive typically tends to make it contextually invariant. The expression tends to pick up the best candidate stable referent across contexts — if there is none, it may end up indeterminate, like “jade”. Thus our linguistically intending to treat “know” as context-insensitive puts pressure in favour

disquotation”, so that in effect my belief is that if it was true to say that I “knew” then, I did “know” that p then, where the quotes in the second occurrences of “know” act as context-shifting operators that hook the word to its semantic value in the earlier context.

Still even if memory does work that way, a speaker who does not notice context shifts will still be likely to believe that if I did “know” that p then, then I did know that p then. By itself, hybrid disquotation devices do not block this kind of error. Something else is needed: either that the speaker is aware of such context shifts, or that the original beliefs and their hybrid disquotation versions should be somehow inferentially encapsulated so as not to be turned into unquoted beliefs.

44. DeRose (2006, §3) now rather argues that our metalinguistic judgements are right: we do take into account context-shifts, provided that we are presented with proper cases that do not involve confusing factors. The proper case are third person ones without retraction or direct disagreement.

45. See DeRose (2006) for an opposite opinion, but Buckwalter (forth.) for some supportive data.
What I point out here is that an analogous point holds for circumstantials and subjunctive conditionals. We have argued that circumstantials give rise to disagreement cases and are fairly resistant to mid-sentence shifts — leaving aside ability and time-sensitive circumstantials (sec. 7.2). Disquotation and retraction cases are not hard to think of. With subjunctive conditionals, the matter is a bit less straightforward. Sobel sequences are not infelicitous; yet they are contradictory if one assumes a context-sensitively strict conditional with a fixed context. This suggests that we tolerate upper shifts of the real possibility standards with conditionals. \(^{46}\) So we would not linguistically intend subjunctive conditionals to be context insensitive. This line of thought can be resisted, though. We may argue that Sobel sequences are minor revisions of one’s claims. A minor revision occurs when one negates one’s earlier claim but reasserts or implies a slightly logically weaker version of the initial claim, as in (7.54).

\[(7.54)\] Everybody was there, or rather, everybody but Mary.

Even though they are contradictory, minor revisions are much smoother to utter than major ones, as (7.55) illustrates.

\[(7.55)\] ?Everybody was there, or rather, not everybody was there.

This suggests that we tolerate upper context shifts of real possibilities. Sobel sequences can be paraphrased as minor revisions (7.56):

\[(7.56)\] All the worlds where Sophie goes to the parade are worlds in which she sees Pedro dance, or rather, all the worlds where Sophie goes to the parade but those in which she is stuck behind someone tall.

A reverse Sobel sequence, by contrast, is akin to the inversion of a minor revision (7.57). Inverted minor revisions are unacceptable, which explains why reverse Sobel sequences are. Consider the following, where the speaker is aware that Marie wasn’t there:

\[(7.57)\] ?Everybody but Marie was there, but everybody was there.

\(^{46}\) The shift is an “upper” one because the sets of real possibilities for the world of evaluation is extended. The new context is stricter for necessity, looser for possibility.
The infelicity of inverted minor revisions is easily explained. There is a point in correcting one’s blunder. But there is none in making the blunder right after the correct claim.

*Revision note.* Timothy Williamson remarks that (7.57) is unfelicitous even if the speaker remembers mid-sentence that Marie *was* there. In such a case, “everybody was there” would be correcting a blunder rather than making one, and so it should be acceptable on my view, but it is not. Now the following explicitly present the same pair as a minor revision, and they are acceptable:

(7.58) Everybody but Marie was there, or rather, everybody was there, even Mary.

This is a minor revision, not an inverted one, and it acceptable. 47 A conditional analogue would be:

(7.59) If Sophie went to the parade but was stuck behind someone tall, she would see Pedro dance, or rather, if she went to the parade, she would see Pedro dance, even if she was stuck behind someone tall. 48

Now the question is why this minor revision cannot be effected by using (7.57). Timothy Williamson suggests the following: for Gricean reasons, the more elaborate claim “everybody but Mary was there” conversationally implies a greater accuracy than the simple “everybody was there”. Thus the latter is not understood as a revision of the former unless it is clearly flagged as such. If that is right, sentences such as (7.57) could only be understood as what I call “inverted minor revision”: namely, trying to

---

47. Perhaps mere stress is sufficient to make it acceptable: “Everybody but Mary was there, or rather, everybody was there.”

48. I should note a disanalogy though: “no” seems to be a good way to make the quantifier version felicitous but not the counterfactual:

Everybody but Mary was there... no, everybody was there.

If Sophie went but was stuck behind someone tall, she would see Pedro dance... no, if Sophie went, she would see Pedro dance.

A temptative explanation: the quantifier version is in fact felicitous only with stress on “everybody”. And in the conditional, there is no word on which one could put the stress.
get away with a slightly-false claim after having uttered the exact truths. And as said, the infelicity of doing that is easily explained.

To sum up: on the minor revision view of Sobel sequences, we do not linguistically intend the conditional to be context-shifting. We simply take the second conditional to be state the truth exactly. The first one is nevertheless regarded as felicitous because it was close to the truth.

A number of phenomena indicate that we linguistically intend “know” to be context-insensitive. Similar phenomena indicate that we linguistically intend modals to be context-insensitive, and perhaps subjunctive conditionals as well. Both phenomena put some pressure in favour of invariantist semantics. On our unified semantics, we expect the decision to go in the same way for “know” and circumstantials.

7.5.3 Knowledge and Action

The links between knowledge and practical reasoning are another source of pressure in the direction of invariantism (Hawthorne, 2004, 86, Williamson, 2005b, §4). The ordinary conception of knowledge ascribes it a central role in practical reasoning.49 For present purposes, we can express it with Hawthorne’s and Stanley’s Knowledge-Action Principle:

**Knowledge-Action principle** Treat the proposition that p as a reason for acting only if you know that p.50

In a contextualist setting, the principle should be conceived as holding at every context. Now, if “know” is sensitive and the knowledge-action principle holds at every context, then what one “should” do is context-sensitive as well (Hawthorne, 2004, 86; Williamson, 2005a, 229). But it is dubious that what one “should” do is a context-sensitive matter. If it is not, given the Knowledge-Action principle, epistemic contextualism is wrong.

---

49. The point has been emphasised by various authors. In addition of the references to Hawthorne and Williamson above, see Unger (1975), Hyman (1999), Williamson (2005a, 227–35), Fantl and McGrath (2002, 2009), Hawthorne and Stanley (2008), Engel (2009).

Contextualism about what one “should” do is dubious because it appears to lose track of the point of “should” claims. A subject $S$ asks: “what should I do?”. On a contextualist view, we may rightfully answer, for some $C, C'$: “$S$ should do $A$” is true in $C$ but “$S$ should not do $A$” is not true in $C'$. To this $S$ is likely to answer: “Right, but what should do I do? Should I do $A$ or not in the end?”. What the answer points to is that we linguistically intend “should” to be context-insensitive. In conjunction with the knowledge-action principle, this is another source of pressure in favour of an invariant semantics for “know”.

I presume that the same will hold for circumstantial modals, though I will not argue the point in detail here. What one should do obviously depends on what could happen. If what “could” happen is a context-sensitive matter, then we expect that what one “should” do will be a context-sensitive matter as well.

### 7.5.4 The price of invariantism

Our unified semantics for modals, subjunctive conditionals and knowledge ascriptions centrally relies on the notion of real possibility (7.3.2). We have initially assumed real possibility to be a context-sensitive matter. Different contexts would count different sets of possibilities as the real ones. The variations would roughly track our variable dispositions to take some possibilities seriously. But the contextualist assumption is inessential to the semantics. It is open to assume instead that there is a unique notion of real possibility — that is, a unique set of real possibilities per world. (One may also consider indeterminacy invariantism, in which the notion is unique but indeterminate.)

In the discussion over contextualism about knowledge ascriptions, many have pointed out that we linguistically intend “know” to be context-sensitive. They argue that this supports an invariantist semantics of “know”. Our intentions put pressure on the word, so to speak, that force it to settle on a unique value.

What we have pointed out is that similar considerations arise in the
case of circumstantial modals. Moreover, given the unified semantics, invariantism about “know” holds only if invariantism about modals does. The semantics is attractive. So it looks as if one has to endorse both invariantism about “know” and about modals or reject both.

This puts the price of invariantism quite high. The orthodoxy is that modals are context-sensitive. Even more strikingly, the invariantist option implies that subjunctive conditionals are strict conditionals — neither variable nor context-sensitive. Many would see the implication as a reductio of the whole view. (I don’t.)

I do not think the issue is easily settled. But I have argued that contextualism about “know” and circumstantial modals stand or fall together. This gives an extremely strong constraint on resolutions of the issue.
Chapter 8

Conclusion

We have defended a theory of knowledge centred around the thesis that knowledge is belief based on an infallible method.

**Method infallibilism**  S knows that $p$ if and only if $S$ believes that $p$ on the basis of a method that could only yield true beliefs.

As we have pointed out in the introduction, the thesis should not be considered as an analysis of the concept of knowledge. The notion of method, in particular, is not given prior to the theory. This gave us two tasks.

First, we had to argue for the thesis. Induction from cases was impossible, since we are not in position to apply the concept of method to cases independently of our applications of the concept of knowledge. Instead, we argued on the basis of structural constraints for any modal accounts of knowledge: the method constraint, the infallibility constraint, and simplicity (ch. 4).

Second, we had to show that, despite the fact that the notion of method is introduced by the thesis, the theory is not trivial. We have done this in three ways. First, constraints on methods are derived from the examination of particular cases and a few natural assumptions (ch. 4, app. E). The same is true of the relevant notion of infallibility (ch. 4, app. D). Second, we make natural and motivated assumptions about their structural properties. The assumptions are explicited in new formal models for
knowledge (ch. 5). Third, we relate the notion to others. Here we have related it to an ordinary notion of real possibility which, we argue, is manifested the semantics of modals and subjunctive conditionals. Methods are related to ordinary possibilities via a semantics of knowledge ascriptions (ch. 7).

Thus developed, the theory reaches into formal epistemology and semantics. The success or failure of such developments is precisely how the theory is tested. Methods earn their keep by their theoretical work, not by their pre-theoretical intuitiveness.

Method infallibilism has also been located in a broader historical landscape (ch. 2). We have put forward a new picture of the history of epistemology. The picture is rough and simple, but it appears much more faithful to history than the widespread legend of the justified true belief analysis. Method infallibilism is located among the family of externalist infallibilist accounts of knowledge. The family is a young one, but history suggests that it is the only one that can avoid both scepticism and the Gettier problem.

In the remaining pages I summarize the results we reached, and I conclude by a consideration of scepticism from the standpoint of the present work.

8.1 History

Chapter 2 defends a New Story of the development of epistemology.

2.1. It is false that Gettier refuted the traditional analysis of knowledge. This Legend appeared in the 1960s. In the 40s and 50s, the traditional view of knowledge was thought be the infallible mental state view defended by Cook Wilson and Prichard.

2.2. The traditional analysis of knowledge was Ur-foundationalism, the view that knowledge consists in a belief bearing a discernible and infallible mark of truth. The discernibility requirement is not essentially internalist but leads to internalism when applied consistently. Ur-foundationalism naturally implies that knowledge is internally accessible and luminous.
2.3. Variants of dogmatic Ur-foundationalism can be found in Aristotle, Epicurus, the Stoics, Descartes, and others.

2.4. *Hume's predicament* is the fact that we do not in fact have infallible and discernible marks of truth for most of our beliefs. Some dogmatic Ur-foundationalists flatly deny it: the Stoics, Locke and Moore. They are mistaken. Ur-foundationalism leads to scepticism: Academic Sceptics, Hume. The *probabilist* tradition is scepticism paired with the idea fallibly justified belief: Carneades, Philo of Larissa, Locke, Peirce, Popper, C. I. Lewis. Idealism and verificationism attempt to recover discernible and infallible marks of truth through revisionary semantics.

2.5. The failure of Ur-foundationalism to avoid scepticism explains the contemporary internalism-externalism divide. Internalism preserves discernibility at the expense of infallibility; externalism preserves infallibility at the expense of discernibility. The Justified True Belief analysis is internalist fallibilism: Malcolm, Ayer and Chishom. It runs into the Gettier problem. Externalist infallibilism avoids the problem by preserving infallibility, but it implies that knowledge is indiscernible. Mixed internalism views try to preserve fallibilist internalism. They face a dilemma: either justification is made epiphenomenal by their anti-Gettier clause or they have secondary Gettier problems. The divide between knowledge and discernibility explains the contemporary focus on justification. The focus covers up a divide between sceptical probabilism and untenable mixed internalism.

### 8.2 Epistemology

Chapter 3 provides a rigorous characterisation and typology of modal requirements on knowledge.

3.1. We introduce a terminology of cases, conditions and requirements. Case are centered worlds with a distinguished proposition. Conditions are sets of cases. Requirements are linguistic or conceptual statements of conditions. Conditions and requirements may be fallibilist or infallibilist. Modal requirements are requirements formulated in terms
of truth, belief, some notion of basis, and some notion of possibility.

3.2. Modal requirements have two (non-decisive) advantages. Infallibility is required to avoid the Gettier problem. Infallibilist requirements need not be modal: undefeasible justification, Plantinga’s proper function, Sosa’s aptness and Greco’s credit, Unger’s non-accidentality. Modal requirements simply state infallibility. Thus they are the more direct way to get an infallible condition. Moreover, modal notions are better understood and formalised.

3.3. Modal requirements require avoidance of error at alternative cases. They are set by four parameters: grounding condition and error type, respect of similarity, and proportion. Infallibility requirements require avoidance of error at all alternative cases, reliability requirements require avoidance of error at most alternative cases. Modal requirements can be proposition-centred, subject-centred, time-centred and variable.

3.4–3.7 review and formalise modal requirements in the literature. We note that: Nozick’s adherence has sceptical consequences and that variable requirements violate deductive closure (3.4.5).

3.8. Summary and discussion of the results. Most requirements are proposition-centred. Goldman’s relevant alternatives requirement and Williamson’s safety may or may not be variable.

Chapter 4 is a defence of method infallibilism. Modal requirements have to be method-based and infallibilist, to avoid the necessary truth problem and Gettier problem, respectively. A simplicity heuristics suggests that they are also sufficient.

4.1. The necessary truth problem: proposition-centred requirements are trivially satisfied by beliefs in necessary truths. As a result, they treat asymmetrically cases that are epistemically symmetrical. They are either an undergeneralisation or made redundant by some stronger condition. Hence the Method Constraint: error should be avoided in cases of belief based on an identical or similar enough method. Methods are just epistemically relevant classes of cases. Epistemically relevant similarity cannot be factored in to component similarities of propositions, bases and other respects. The Method constraint can be satisfied by method-identity
requirements and *method-similarity* requirements. There are considerable reasons to favour the second type. But for simplicity, we focus on the first.

4.2 Fallibilism grants knowledge in lottery cases, but infallibilism denies knowledge in ordinary knowledge case. To avoid the dilemma, one should consider the structure in ordinary cases, i.e. a difference between favourable and unfavourable areas of cases. Infallibility in a favourable area is required, on pain of Gettier cases. We represent the argument formally. From this and the Method constraint, method infallibility is necessary for knowledge.

4.3 Two objections to necessity: the generality problem and the sceptical problem. The generality problem is the demand for a definition of methods in epistemically or metaphysically prior terms. The demand is unwarranted, though it is motivated by two demands that are reasonable: that the account be testable and reductive. The concept of knowledge is primitive: it is not analysable in terms of more basic ordinary concepts. We do not have an adequate pre-theoretical conception of methods. Methods are picked up by their role in the method infallibilist of knowledge. The account is tested by our ability to find a natural enough notion of method. The account can be ontologically reductive. The strategy, thus, is to use intuitions about knowledge to find out what methods are. We apply it to a few cases, and conclude that methods can be very fine-grained, satisfy a limited inernalist constraint, and are not luminous. The sceptical problem is that requiring infallibility appears to lead to scepticism. Here as well, one should use intuitions about knowledge to determine what methods and infallibility are. The strategy is applied to three types of sceptical arguments: Evil-Demon scenarios, quantum-mechanical lotteries and ordinary possibilities of error. Insofar as there is an intuition that one knows in such cases, it can be matched with a reasonable individuation of methods and possibility such that error cases are relevantly different from knowledge cases. When no reasonable difference is found, the intuition that subjects know vanishes. Method infallibilism thus adverts sceptical consequences in cases in which they
are clearly unacceptable. But it may not advert them everywhere, which shows that the strategy is not trivial.

4.4. Three objections to the sufficiency of method infallibility: the need for internal justification, the absence of defeaters, and reliability or virtue; and the Simplicity heuristics in favour of sufficiency. Internal justification is not needed. The Clairvoyant case involves a fallible method. Only the absence of real defeaters is needed, but real defeaters are just those that show one’s methods to be fallible. Adding a virtue or reliability requirement leads to three problems: a worse generality problem, secondary Gettier cases and the necessity to distinguish reliable methods that are knowledge-apt from those that are not. The cases that motivate the additional requirement are shown to involve fallible methods once we remember that methods are not tied to a particular belief or a restricted set of propositions. In favour of sufficiency, a Simplicity heuristics is put forward. Accounts of knowledge made up of independent conjunctions are liable to secondary Gettier problems. When faced with counterexamples to sufficiency, one should not add conditions to one’s account, but replace one’s condition with a stronger one. This puts the bar high for competitors of the method infallibility requirement, and provides defeasible support for its sufficiency.

8.3 Formal epistemology

Chapter 5 provides a formal representation of methods and method-based models of belief and knowledge.

5.1. Knowledge is uncontroversially a matter of the basis of one’s belief. Standard Hintikka-style models do not provide a natural way to formalise the idea. They represent knowledge as the elimination of possibilities, where a possibility is eliminated if incompatible with what one knows. This implies logical omniscience and is at odds with inductive knowledge. Instead we represent knowledge as belief based on an infallible method.
5.2. Methods are defined as functions from worlds and sets of premises to sets of conclusions. Method union and method combination are defined. A background alethic modality is introduced to define infallibility.

5.3. Methods-based representations of various epistemological notions: Gettier cases, Gettier cases involving a belief in a necessary truth, inductive knowledge, fallible but reliable methods.

5.4. Methods-based models. We formalise propositions as set of worlds. Frege cases cannot be dealt with, but the models are kept simple and are shown to be equivalent to neighbourhood models (app. C.2). We assume agents with no bounds on method union and combination.

5.5. In general, methods-based models validate the idea that knowledge entails true belief. However, the full S5 system is derived with three sets of natural idealisations of agents. Deduction and Pure Reason methods validate K, or perfect reasoning. Introspection and Introspective Confidence methods validate knowledge of one’s beliefs and a limited form of negative introspection, p5, according to which one knows that one does not know what one fails to believe. 4, or positive epistemic introspection, is derived as well, provided that the background alethic modality is transitive. Excellence is not a method, but the fact that all the agent’s methods are infallible. Excellence implies the equation of belief and knowledge. In conjunction with Introspective confidence, it entails 5, or negative epistemic introspection. The results provide an explanation of why and when the S5 axioms hold, and a vindication of their use. They suggest that the S5 axioms do not embody an ideal of subjective rationality, but three different ideals: pure rationality, excellence with respect to the inner, excellence with respect to the outer.

Chapter 6 presents and discusses DeRose’s and Lewis’s modal versions of epistemic contextualism.

6.1. Contextualism about “know” is the view that “know” expresses different conditions at relevantly different contexts. Modal epistemic contextualists take the different conditions expressed by “know” to vary along different sets of alternative cases.

6.2. Lewis’s takes “know” to be sensitive to a set of relevant possible worlds provided by the context. The set is determined by a set of Rules, some of which introduce or exclude worlds in virtue of the speaker’s situation, others in virtue of the subject’s situation. Yet even the later are speaker-sensitive, since their effect depends on who the speaker discusses. The Rules are specific to “know”. Lewis’s notion of possibilities is problematic: they need to be specific enough not to systematically result in scepticism, but unspecific enough so that one can attend to them. The problem is avoided if attending to a possibility brings in the closest possible worlds in which it is true.

6.3. DeRose takes “know” to be sensitive to distance measure between worlds. DeRose’s idea that methods are part of the closeness relation is untenable. Context-shifts are governed by the Rule of Sensitivity but not automatic. The Rule is specific to know. The Rule has no analogue elsewhere in language.

6.4. DeRose’s and Lewis’s view are proposition-centred and consequently face the necessary truths problem. Lewis’ linguistic guise answer is untenable. Blome-Tillmann’s proper basing answer faces a range of issues. Again, methods should be introduced.

6.5. Attending to some possibilities is not sufficient to make them relevant to “know” attributions. Most agree that possibilities of error generate sceptical intuitions only if they are taken seriously, but an account of taking seriously is lacking. On the no-belief account of taking seriously, a possibility is taken seriously by one iff one does not believe it to be false. The proposal, and Blome-Tillmann’s related presupposition
6.6. Lewis’s account violates the factivity of knowledge for unmentioned subjects, because his semantics fixes a set of possibilities directly as a function of the context. DeRose’s view avoids the problem, by having the context selecting an accessibility function that in turn selects possibilities. However, this feature of DeRose’s account, combined with his Rule of Sensitivity, has the puzzling consequence that sceptical scenarios involving other-worldly subjects are not efficient in raising sceptical intuitions.

Chapter 7 proposes a unified semantics of modals, knowledge and subjunctive conditionals in terms of real possibility. The semantics can be interpreted as contextualist or invariantist.

7.1. We give an overview of modal auxiliaries based on Kratzer’s and Hacquard’s work. The modals that interest us are the ones that express an alethic modality: circumstantial. We reject Hacquard’s suggestion that circumstantial are substantially event-relative.

7.2. Some circumstantial modals are reluctant to having multiple values in a single context. We propose a contextual parameter, real possibility, that takes a unique value for a given context. We set aside circumstantial modals with an ability reading. Time-relative circumstantial modals are themselves dependent of real possibility.

7.3. An attractive semantics of subjunctive conditionals also relies on a single real possibility parameter of the context. We examine some rules for the dynamics of real possibility. A context may mention a possibility as a non-real one; we call them theoretical possibilities.

7.4. A semantics for “know” in terms of real possibilities is given. At every context, “know” requires avoidance of “real” possibilities of error. Taking seriously a possibility believing it to be part of the real possibilities. The account has a number of advantages over Lewis’s and DeRose’s: factivity, non-automaticity, no use of the Rule of Sensitivity, an account of third-person sceptical arguments, and no need to postulate a semantic exception. The account predicts links between knowledge ascriptions and conditionals more accurately than Sensitivity and (conditional-based)
Safety.

7.5. Three considerations appear to put pressure in favour of an invariantist account of “know”: naturalness, semantic blindness and the link between knowledge and action. The first can be met by a contextualist account. But the second and the third do give reasons to think that “know” would pick up, as far as possible, a unique condition for all contexts. We point out that both considerations apply equally to modals and subjunctive conditionals. In conjunction with the unified semantics, this imply that contextualist about “know” and “modals” stand or fall together. The price for invariantism about “know”, namely a context-insensitive strict-conditional analysis of subjunctive conditionals, is high.

8.5 The two sources of scepticism

Our work identifies two sources of scepticism.

The first is the elusiveness of methods. Knowledge requires a method that avoids error. However, we do not have an ordinary conception of methods. The history of epistemology suggests that if anything, our heuristics for knowledge ascription lead us to assume that methods should be discernible. This leads to internalist conceptions of evidence and justification, and to Cartesian skepticism.

The second is the elusiveness of possibility. Knowledge requires a method that avoids real possibilities of error. But the notion of real possibility is elusive. Our conception of it is shifty and may indicate that real possibility is context-sensitive. We are easily tempted to allow all physical possibilities to be real. This leads to sceptical arguments that do not rely on an internalist conception of methods, such as Quantum-mechanical scepticism.

Arguments of the first kind are not set aside simply by rejecting internalism. Even with an externalist individuation of methods, it is an open question whether natural ways to individuate methods do vindicate the idea that many of our methods are infallible. Arguments of the second kind are not definitely set aside if we take real possibility to be context-
sensitive. Even if real possibility was largely determined by what we take seriously, it would still not be equated with what we take seriously. Our failure to notice tiny but close possibilities of error may not be sufficient to discard them. We may be safe from Cartesian scepticism, but not safe from a moderate but widespread form of scepticism in which all our methods are at most quasi-infallible. This turns out to depend on the fine details of method individuation and the semantics of real possibility. The sceptical problem is not solved.
Part V
Appendices
This appendix to chapter 3 reviews semantics of subjunctive conditionals that can be put to use in formulating modal requirements on knowledge. Sections A.1-A.4 give an overview of the semantics. In the main text (ch. 3, sec. 3.4.4 and 3.5.3) we have given “quasi-Nozickian” semantics for Nozick’s tracking conditions under the Limit Assumption. In section A.5 below we give the Nozickian versions and the Quasi-Nozickian without the Limit Assumption.

A.1 Semantics for subjunctive conditionals

The rough idea of modal semantics for subjunctive conditionals is that a subjunctive conditional if $p$ would $q$ ($p \rightarrow q$) is true at a world $w$ iff the material conditional $p \rightarrow q$ holds throughout one or several relevant spheres of worlds around $w$. More precisely, the semantics can be singular or quantified. On the simple semantics, an utterance of the subjunctive conditional $p \rightarrow q$ will be true iff the material conditional $p \rightarrow q$ holds within a given relevant sphere of worlds. On the quantified semantics, we rather specify conditions that a sphere must meet to be relevant, and we say that the utterance is true iff $p \rightarrow q$ holds in some (existential version) or all (universal version) relevant spheres. The singular semantics is a degenerate case of the quantified one with a unique relevant sphere.
Moreover, a universal semantics can be turned into a singular one where the single relevant sphere is the union of all the relevant ones. Existential semantics, however, are essentially quantified.

These semantics can be further put into two broad kinds: *variably strict* semantics and *context-sensitively strict* semantics. The two differ on how they take relevant spheres to be selected.

The *variably strict conditional* approach is that of Stalnaker (1968) and Lewis (1973, chap.1), which Nozick follows. On this kind of view, the *antecedent proposition* determines the conditions on relevant spheres. Different sets of spheres are considered for different antecedents. Thus it straightforwardly follows that conditionals do not contrapose: it is not the case that \( p \rightarrow q \) iff \( \neg q \rightarrow \neg p \), for \( p \) and \( \neg q \) typically send us to different sets of relevant spheres.

The *context-sensitively strict conditional* approach is that of von Fintel (2001) and Gillies (2007). (It was considered in Lewis (1973, 13) and defended by Warmbrod (1981a,b); see also Wright (1983, 137–8).) On this kind of view, the *context* determines the relevant spheres. The spheres thus remain *fixed* from one conditional to another, provided context does not change. On the single and universal semantics, it straightforwardly follows that conditionals contrapose: in any given context, \( p \rightarrow q \) iff \( \neg q \rightarrow \neg p \), for both require \( p \rightarrow q \) to be true at all relevant spheres of the context. If the conditionals do not appear equivalent, it is only because uttering them affects the context.

A *hybrid* semantics is considered by Williamson (2000, 149) (who calls it “Nozickian”) and perhaps assumed by Sosa (1996) in his subjunctive conditional formulation of safety. On this hybrid view, the relevant spheres depend on the antecedent proposition when the antecedent proposition is *false*, and depends on a fixed threshold when the antecedent proposition is *true*.  

---

1. The label “variably strict” comes from Lewis (1973, I, §1.3). Not all existing semantics or accounts of subjunctive conditionals fall into these kinds, but the ones we discuss here do. They include the most commonly used by philosophers. Kratzer (2010a, 28) argues that conditionals should not be regarded as two-place propositional connectives, contrary to what philosophers commonly do; but her semantics for subjunctive conditionals (Kratzer, 2010b, 45) is truth-conditionally equivalent to that of Lewis (1973, 16).
true.

Mixed semantics are those for which relevant spheres are selected on the basis of both the antecedent proposition and the context. Lewis’ (1973, 67) original semantics is in fact mixed, since he takes context to resolve vagueness in the ordering based on which the antecedent proposition determines relevant spheres. For most intents and purposes, mixed semantics can be grouped with variably strict ones, for their logics for conditionals have the same features (e.g. failure of contraposition).

As a rule of thumb, requirements on knowledge that rely on variable sets of spheres are incompatible with deductive closure. If the spheres relevant to evaluating some condition on knowing \( p \) differs from the spheres relevant to evaluating some condition on knowing \( q \), then the fact that \( p \) entails \( q \) will typically not guarantee that the first condition entails the second. A prime example is Nozick’s sensitivity condition. But Nozick’s adherence violates closure as well, and so does safety on a variably strict semantics. One needs always to check the details (notably the role of “methods” or bases of knowing) of a particular account, but as a rule of thumb, accounts relying on conditionals with variably strict semantics are expected to be incompatible with closure.

All the semantics rely on a similarity ordering over worlds. We introduce a three-place relation \( w' \) is closer to \( w \) than \( w'' \), notated \( w' \preceq_w w'' \), such that \( \preceq_w \) is a complete preorder over a set of worlds \( S_w \) with \( w \) as a minimal element. A preorder is a reflexive and transitive relation: for any \( w' \) in \( S_w \), \( w' \preceq_w w' \), and for any \( w', w'', w''' \) in \( S_w \), \( (w' \preceq_w w'' \wedge w'' \preceq_w w''') \rightarrow w' \preceq_w w''' \). The preorder is complete over \( S_w \) iff for any \( w', w'' \) in \( S_w \), \( (w' \preceq_w w'' \vee w'' \preceq_w w') \). \( w \) is a minimal element iff for all \( w' \) in \( S_w \), \( w \preceq_w w' \). Worlds \( w^* \) out of the set are such that neither \( w^* \preceq_w w' \) nor \( w' \preceq_w w^* \) holds for any \( w' \); they are called inaccessible from \( w \). From the similarity order we derive a system of nested spheres \( S_w^* \): each sphere \( S_{w,w'} \) in \( S_w \) contains all the worlds that are as close to \( w \) as some world \( w' \) in \( S_w \). The order guarantees that the spheres are nested: for any two spheres in the system, one contains the other. \( S_{w,w} \) is the smallest sphere in \( S_w \): the set of worlds that are as close to \( w \) as \( w \) itself. The accessibility sphere of
$w$ is the intersection of all the spheres in the system: $S_w = \bigcap S_w$.

We say that a sphere permits $p$ iff $p$ holds in some world in the sphere and that a sphere enforces $p$ when $p$ holds throughout the sphere.

Additional assumptions are often made, e.g. that $w$ is the unique closest world to itself, or that for any $p, w$, there is a unique closest $p$ world to $w$. In chapter 3, we have relied on the Limit assumption: for any antecedent $p$, if there is a $p$-permitting sphere, then there is a smallest $p$-permitting sphere. This is not guaranteed in general, since it may happen that for any $p$ world, there is some $p$ world closer than it, and so on indefinitely (Lewis, 1973, 19-21). The assumption allowed us to reduce quantified semantics such as Lewis’s to singular ones and simplify the presentation. Here we will not make the assumption and formulate the semantics more generally. As a result, we will have quantified modal requirements, distinct from the singular ones we formulated in the main text.

We formulate all semantics as variations on the following patterns:

**Universal** $p \rightarrow q$ is true at world $w$ and context $c$ iff $p \rightarrow q$ holds throughout all spheres in $S_w$ satisfying condition $F(p, c)$.

**Existential** $p \rightarrow q$ is true at world $w$ and context $c$ iff (a) $p \rightarrow q$ holds throughout some sphere in $S_w$ satisfying condition $F(p, c)$, or (b) $p \rightarrow q$ holds at all worlds in $\bigcap S_w$.

Clause (b) of the existential version is meant to deal with cases in which no sphere satisfies $F(p, c)$. It can be dropped if we ensure that $\bigcap S_w \in S_w$ and that $\bigcap S_w$ satisfies $F(p, c)$ for any $p, c, w$.

In variably strict semantics, the condition $F(p, c)$ is a function of the antecedent $p$ only; in context-sensitively strict semantics it is a function of context $c$ only.

Henceforth I will simply speak of “conditionals”, but only subjunctive ones are intended.

---

2. See Lewis (1973, 14-5) on the system of nested spheres and Lewis (1973, 48-50) on its equivalence to a preorder over a subset of worlds.
A.2 Variably strict semantics

A.2.1 Lewis's semantics

Lewis's (1973, ch. 1) semantics is:

\[ p \rightarrow q \] is true at \( w \) iff \( p \rightarrow q \) holds throughout some \( p \)-permitting sphere in \( S_w \), or there is no \( p \)-permitting sphere in \( S_w \).

That is, the relevant spheres are the \( p \)-permitting ones, and we existentially quantify over them. By nesting, if there is a smallest \( p \)-permitting sphere, the semantics reduces to: \( p \rightarrow q \) holds throughout the smallest \( p \)-permitting sphere. If there is no smallest \( p \)-permitting sphere, then \( p \rightarrow q \) should hold throughout some \( p \)-permitting sphere, and by nesting, throughout all the spheres smaller than it. When the antecedent is true, there is a smallest \( p \)-permitting sphere, namely the set of all worlds that are as close to actuality as actuality itself. If the actual world is the unique closest world to itself, the conditional comes out true simply if both antecedent and consequent are true.

A.2.2 Lewis-Nozick semantics

Nozick finds Lewis's semantics too weak for conditionals with true antecedents. He suggests that for \( p \rightarrow q \) to be true, \( q \) must be true in all \( p \) worlds up to the first \( \neg p \) ones. More precisely, in all the worlds strictly closer than the first \( \neg p \) ones (Nozick, 1981, 680n). By nesting, this means all the spheres that include \( p \) worlds only; if there is no such sphere, then presumably we require \( p \rightarrow q \) to hold throughout the smallest sphere. Combining this idea with Lewis's semantics for the conditionals with false antecedents, we get:

\[ p \rightarrow q \] is true at \( w \) iff:

- If \( p \) is true at \( w \), \( p \rightarrow q \) holds throughout \( S_{w,w} \) and all \( p \)-enforcing spheres in \( S_w \).
- If \( p \) is false at \( w \), either \( p \rightarrow q \) holds throughout some \( p \)-permitting sphere in \( S_w \), or there is no \( p \)-permitting sphere in \( S_w \).
Say that a proposition is a necessary truth at \( w \) if all spheres in \( S_w \) enforce it.\(^3\) The difference with Lewis’s semantics is dramatic when \( p \) is a necessary truth. On Lewis’s semantics, if \( p \) is a necessary truth, \( p \rightarrow q \) is true if \( p \rightarrow q \) holds in the smallest sphere — so if the actual world is the unique closest to itself, the actual truth of \( q \) is sufficient. On the Lewis-Nozick’s semantics, if \( p \) is a necessary truth, \( p \rightarrow q \) is true only if \( q \) holds throughout all accessible worlds, that is, if \( q \) is necessarily true. (Unfortunately intuitions are of no help to adjudicate.)

The Lewis-Nozick semantics is asymmetric: not only the relevant spheres are different for true and false antecedents, but the truth-condition for true antecedents universally quantifies over relevant spheres while the one for false antecedents existentially quantifies over them. Note in particular that when the antecedent is true, we go up to but excluding the closest spheres in which \( \neg p \) holds. Why would that matter, since \( p \rightarrow q \) trivially holds at \( \neg p \) worlds? Because those spheres may also contain further \( p \) worlds that are just as close as the closest \( \neg p \) ones. According to Nozick’s truth-conditions for true antecedents, where \( q \) is also true in those worlds does not matter. That is why the relevant spheres are only the ones before the closest \( \neg p \)-permitting spheres.

### A.2.3 Quasi-Nozickian semantics

Thus a noteworthy alternative is one in which the relevant spheres go up to and including some sphere including a \( \neg p \) world. Combined with Lewis’s semantics for the false antecedent case, we get a more unified semantics:

**Quasi-Nozickian semantics** \( p \rightarrow q \) is true at \( w \) iff: either (a) \( p \rightarrow q \) holds throughout some sphere in \( S_w \) that permits both \( p \) and \( \neg p \), or (b) \( p \rightarrow q \) holds at all worlds in \( \cap S_w \).

Here the relevant spheres include all spheres in which some world differs from actuality on the truth-value of \( p \): the first \( p \) spheres (and any bigger sphere) when \( p \) is false, and the first \( \neg p \) spheres when \( p \) is true. If there

---

3. This is what (Lewis, 1973, 22) calls “outer necessity”.

is no such sphere, then either \( p \) fails everywhere and the conditional is vacuously true, or \( p \) holds everywhere and \( q \) must hold everywhere as well. The semantics is slightly stronger than the Lewis-Nozick one.

The semantics has some motivation in the context of Nozick’s analysis of knowledge. The idea is to evaluate counterfactuals with a \( p \) or \( \neg p \) antecedent over sets of worlds inclusive enough to have both \( p \) and \( \neg p \) worlds. This gives a background for evaluating whether belief is non-trivially responsive to differences in truth-value.

### A.2.4 Nozick’s Stretch semantics

Nozick (1981, 680–1n) himself thought that the Lewisean truth condition is also too weak for conditionals with false antecedents. On his view, when \( p \) is false, \( q \) should not only be true at the closest \( p \) worlds, but also “further out”. He sketches two stronger semantics: the *Stretch* semantics and the *Weakened Spread* semantics.

The *Stretch* semantics extends his truth-condition for conditionals with true antecedents. The idea is that if at the closest \( p \) worlds, we encounter a pure stretch of \( p \) worlds, we should require \( q \) to hold as far as the stretch extends. Say that a \( p \) world \( w' \) is separated from a strictly closer \( p \) world \( w \) iff there is a \( \neg p \) world \( w^\emptyset \) between them: \( w \preceq w^\emptyset \prec w' \). Nozick introduces the notion of \( p \)-neighbourhood as the set of worlds that are not separated from a closer \( p \) world, and says that \( p \rightarrow q \) is true at \( w \) iff \( q \) holds throughout the \( p \)-neighbourhood of \( w \).

**Nozick’s stretch semantics** \( p \rightarrow q \) is true at \( w \) iff \( p \rightarrow q \) holds at all spheres in \( S_w \) such that no \( p \) world in the sphere is separated from any closer \( p \) world.

That is, the relevant spheres are all spheres whose \( p \) worlds all belong to

---

4. “The \( p \) neighbourhood of a world is the closest \( p \)-band to it, that is, \( w \) is in the \( p \) neighbourhood of the actual world if and only if \( p \) is true in \( w \) and there are no worlds \( w^\emptyset \) and \( w^\emptyset' \) such that not-\( p \) is true in \( w^\emptyset \) and \( p \) is true in \( w^\emptyset' \), and \( w^\emptyset ' \) is closer to \( A \) than \( w \) is to \( A \), and \( w^\emptyset \) is at least as close to \( A \) as \( w^\emptyset ' \) is to \( A \). A subjunctive \( p \rightarrow q \) is true iff \( q \) holds throughout the \( p \) neighbourhood of the actual world.” (Nozick, 1981, 680n) Nozick’s use of “band” here differ from ours below.
the \( p \) neighbourhood. To spell out the consequences, we will call a \textit{band} a set of worlds that are equally close. A \textit{pure} \( p \) \textit{band} is a band throughout which \( p \) holds; and impure \( p \) band is a band containing both \( p \) worlds and \( \neg p \) worlds.\(^5\)

If there are closest \( p \)-worlds, the outcome is straightforward: the \( p \) neighbourhood contains all \( p \) worlds that are not separated from the closest ones. This means all the closest \( p \) worlds; furthermore, if the closest \( p \) worlds are in a pure \( p \) band, all \( p \) bands contiguous to it; furthermore, if there is a first impure \( p \) band contiguous to the pure ones, the \( p \) worlds of that band.\(^6\) If \( p \) is true, the actual world is a closest \( p \) world, so we are in one of these cases.

If \( p \) is false, we may have no closest \( p \) world, but an infinitely descending series of closer and closer \( p \) worlds. Either such a series starts with a pure stretch, or it does not — that is, either there is some world \( w \) such that any closer \( p \) world is in a pure \( p \) band, or there is not. If it does, the \( p \) neighbourhood includes the initial pure \( p \) bands and any contiguous ones; furthermore, the \( p \) worlds of a first and contiguous impure \( p \) band, if any. But if we do not have an initial pure stretch, then every \( p \) world has a closer \( p \) world within an impure band, and is therefore separated from it. As a result, the \( p \) neighbourhood is empty and the conditional comes out as true no matter what the consequent is, but it should not (Nozick, 1981, 681n).\(^7\)

---

5. Nozick also uses the term “band”, but to refer to stretches (1981, 680n) or to spreads (1981, 681n).

6. If there is a band of closest \( p \)-worlds but it is impure, then all further \( p \) worlds are separated from the closest by the \( \neg p \) worlds in that band. If a \( p \) world is in the first impure band contiguous to an initial pure stretch of \( p \) bands, then any \( \neg p \) world that is \textit{strictly} closer than it will be before the initial \( p \) stretch itself; so there will not be a \( p \) world at least as close as that \( \neg p \) world; so our world will be included in the \( p \) neighbourhood. If after an initial stretch of pure \( p \) bands, we have an infinitely descending series of smaller and smaller spheres with impure bands, then any \( p \) world in those spheres will be separated from the pure \( p \) bands by the \( \neg p \) world of some smaller sphere.

7. A tentative illustration: it is random whether a photon fired from a source would go through the right slit or the left one. So presumably, for any world where the photon goes through the right, there is a world just as close where it goes to the left. Suppose additionally that there are no closest worlds in which the photon is fired: say that the photon would be fired if the source was heated up for any time strictly greater than
A.2. Variably strict semantics

For conditionals with true antecedents, the truth-condition will be equivalent to the Lewis-Nozick one, except when there is an initial pure stretch followed by a first and contiguous impure $p$ band: then the first impure $p$ band is included and the truth-condition is equivalent to the stronger Quasi-Nozickian one. For conditionals with false antecedents, when there are closest $p$ worlds, the truth-condition is at least as strong as the Lewisean one — equivalent to it if the closest $p$ band is impure, stronger if it is pure and followed by further pure ones or a first and contiguous impure one.

To deal with the case in which there is no closest $p$ worlds and no initial pure stretch, we may stipulate that the truth-condition for that specific case is the Lewisean one. The resulting semantic is a baroque one in which a universally quantified truth condition becomes existentially quantified in that specific case.

A.2.5 Strict Spread semantics

Nozick instead proposes an alternative, the Weakened Spread semantics. He introduces it with two moves: one that strengthens the stretch semantics, and a second one that weakens the results of the first.

A stretch is a series of $p$ worlds that are not separated by an intervening $\neg p$ world. A spread is a series of $p$ worlds that are all connected by intervening $p$ worlds. Say that a $p$ world $w'$ is connected to a strictly closer $p$ world $w$ iff any $\neg p$ world between them is matched by some $p$-world: for any $\neg p$ world $w_\phi$ such that $w \leq w_\phi < w'$ there is a $p$ world $w_\theta$ such that $w_\theta \leq w_\phi$ and $w_\phi \leq w_\theta$. A pure $p$ spread is a set of connected $p$ worlds. The analogue of the stretch semantics is:

---

2 seconds, so that (we assume) for any world in which the source was heated up for $2 + k$ seconds, there is a closer world in which the source has been heated up for $2 + k/2$ seconds. The conditional “If the electron went through the right, the world would end” may be false, but the semantics will automatically count it as true.

8. “let us redefine the relevant $p$ band as the closest spread of $p$ worlds such that there is no not-$p$ world intermediate in distance from $A$ to two $p$ worlds in the spread unless there is also another $p$ world in the spread the very same distance from $A.”$ (Nozick, 1981, 681n)
**Strict Spread semantics** \( p \rightarrow q \) is true at \( w \) iff \( p \rightarrow q \) holds at all spheres in \( S_e \) such that all \( p \) worlds in the sphere are connected to any closer \( p \) worlds.

Equivalently, at all spheres whose \( p \) worlds are all within the closest \( p \) spread. As opposed to \( p \) neighbourhoods, the existence of a closest \( p \) spread is guaranteed when there is some accessible \( p \) world. So the semantics avoid the trivialization that the stretch semantics encounters.

If a world is not separated, it is connected. Hence the semantics is stronger than the Stretch Semantics.\(^9\) In fact, it is the strongest of the ones we have seen so far: a counterfactual evaluated as true by this semantics will be true by the Lewisean, Lewis-Nozick, and Quasi-Nozickian semantics as well.

Nozicks (1981, 681n) finds the semantics ill motivated. The closest \( p \) spread contains the first \( p \) band, be it pure or impure, — or the initial \( p \) bands if there is no first one — and any \( p \) band contiguous to it, and so until some pure \( \neg p \) band is encountered. Why stop at a pure \( \neg p \) band? It is not clear what is special about it.

### A.2.6 Nozick’s Weakened Spread semantics

Nozick does not endorse the strict spread semantics, but gestures towards a weakened version of it:

Perhaps nothing stronger can be said than this: \( p \rightarrow q \) when \( q \) holds for some distance out in the closest \( p \) band \([=\text{spread}]\) of the actual world, that is, when all the worlds in this first part of the closest \( p \)-band \([=\text{spread}]\) are \( q \). The distance need not be fixed as the same for all subjunctives although various general formulas might be imagined, for example, that the distance is a fixed percentage of the width of the \( p \) band \([=\text{spread}]\).

(Nozick, 1981, 681n)

Thus we would have something along the lines of:

---

\(^9\) In the particular cases in which the Stretch Semantics is trivialized, it counts the counterfactuals as true, so the Spread semantics can only be stronger.
Nozick’s Weakened Spread semantics  We assume a function that assigns to each spread of worlds an initial part of the spread.

\[ p \rightarrow q \text{ is true at } w \text{ iff } p \rightarrow q \text{ holds at all spheres in } S_w \text{ such that all } p \text{ worlds in the sphere are in the initial part of the closest } p \text{ spread.} \]

This would be a crucial weakening of the semantics. When \( p \) is true, it only requires that \( p \rightarrow q \) holds some distance out; not necessarily up to \( \neg p \)-worlds. Thus the truth-condition for a conditional with a true antecedent may be weaker than in the Lewis-Nozick, Quasi-Nozickian and Stretch semantics. If so, it may avoid the (overlooked) sceptical consequences of Nozick’s adherence condition (see ch. 3 sec. 3.4.5). But it may also be stronger, for instance if the closest \( p \) spread extends to all accessible worlds. That will depend on the modal layout and on how the “initial part” of the spread is defined.

Similarly, when \( p \) is false, the semantics requires that \( p \rightarrow q \) holds for some distance out after the first \( p \) worlds, but again, how far out will depend on the size of the spread. So again, the truth-condition may be weaker or stronger than the Lewis-Nozick, Quasi-Nozickian and Stretch semantics.

Summing up: the weakest semantics is the Lewisean and the strongest one the Strict Spread.\(^{10}\) The Lewis-Nozick one is stronger than the Lewisean. Then come the Quasi-Nozickian and Stretch semantics, ordered as follows: the Quasi-Nozickian is stronger for conditionals with true antecedents, but the Stretch semantics is stronger for conditional with false antecedents.\(^{11}\) The Weakened Spread semantics lies between the Lewisean and the Strict Spread ones as well, but cannot be ordered respectively to the others.

\(^{10}\) Assuming that we patch the Stretch semantics with e.g. a Lewisean truth condition in cases in which it is trivialized. Otherwise it will be weaker than the Lewisean ones in those cases.

\(^{11}\) Again, assumed the Stretch semantics is patched.
A.3 Context-sensitively strict semantics

The context-sensitively strict semantics is simply this (von Fintel, 2001, 138; Gillies, 2007, 337, 351):

**Context-sensitively strict semantics** \( p \models q \) is true in context \( c \) at \( w \) iff \( p \rightarrow q \) holds throughout \( S_{w,c} \), where \( S_{w,c} \) is the sphere fixed by context \( c \) for world \( w \).

Stated thus, the relevant sphere could be any arbitrary selection of worlds. von Fintel (2001, 135) and Gillies (2007, 336) require spheres to *nested* with respect to a Lewis-style closeness ordering:

**Nesting** For any \( S, c, w, w', w'' \), if \( w'' \in S_{w,c} \) and \( w' \preceq w'' \), then \( w' \in S_{w,c} \).

Moreover, von Fintel (2001, 137) and Gillies (2007, 333) assume that an *utterance* of a conditional \( p \models q \) typically affects the context in such a way that the sphere of the context includes some \( p \) worlds, if there are any. On von Fintel’s view this is the result of a presupposition on the part of the conditional (von Fintel, 2001, 124). Gillies (2007, 344) rejects the presupposition account, and argues instead that a conditional in effect carries a claim that the antecedent is possible (Gillies, 2007, 349). In any case the result is:

**Entertainability constraint** For any context \( c \) in which a conditional “\( p \models q \)” is successfully uttered, \( S_{c,w} \) includes some \( p \) worlds for any \( w \).

(The constraint does not cover cases in which \( p \) is contradictory or impossible. If needed, it can be amended so that conditionals with impossible antecedents are trivially successful.)

Finally, the entertainability constraint is applied in a conservative manner: when “\( p \models q \)” is uttered at a context whose spheres do not include \( p \) worlds, the context is modified so as to expand these spheres up to to the *closest* \( p \) worlds:

**Dynamic update** If a conditional “\( p \models q \)” is uttered at a context \( c \) such that there are no \( p \) worlds in some \( S_{c,w} \), then if possible, the context is changed to a new context \( c' \) in which for any \( w, S_{c',w} = S_{c,w} \cup \{w' :
A.3. Context-sensitively strict semantics

∀w''(w'' ⊨ p → w' ≤ w''}, and the counterfactual is evaluated at c∗.

Both von Fintel (2001, 145n4) and Gillies (2007, 335n8) rely on the Limit Assumption, which holds if the set of worlds is finite but may fail in infinite settings. It is noticeable that avoiding the assumption would significantly change the picture of dynamic update. Suppose we have an infinite series of closer and closer p worlds, all of which are out the sphere of the current context. A conditional “p □→ q” is uttered: the sphere should be expanded so as to include p worlds, but up to where? On the Lewis view, the counterfactual should come out true as long as there is some p world such that p → q holds throughout all worlds at least as close as that one. So if there is some such world, the sphere should only be expanded up to that world but no further. This would mean that the update also depends on the consequent. If there is no such world, then presumably some arbitrary small p-permitting sphere should be picked up. Alternatively, one could claim that the update picks up some arbitrary but reasonably small p-permitting sphere; thus the truth-conditions would be sometimes stricter than the Lewis ones. Both solutions appear partly ad hoc. A more principled one would be to introduce a system of spheres as a contextual parameter, instead of a single one, and to use the Lewis truth-condition (namely, p □→ q holds only if p → q holds in some sphere in the system, or no sphere permits p). We start with a limited system, e.g. one such that no sphere permits p; when encountering a “p □→ q” conditional, we expand the system to include some of the closest p-permitting spheres (if there are any). But here we will follow von Fintel and Gillies and assume that there are always closest p worlds.

How close is the contextualist package to Lewis’s semantics? Suppose “p □→ q” is uttered at c, w and the entertainability constraint is satisfied. Since p is entertainable, there are p-worlds in Sw,c. Since the sphere is nested, at least the closest p-worlds are included. So Sw,c is at least as big as the Lewis sphere for w and p. But it may include more. So when the conditional is uttered and the entertainability constraint is satisfied, the von Fintel-Gillies conditional entails the Lewis one. When the conditional is
not uttered or the constraint not satisfied, all the bets are off. For instance, suppose that $S_w$ does not include $p$ worlds: the closest $p$-worlds, in particular, are out of the sphere. Then the non-uttered conditional $p \not\rightarrow q$ is true in the context, even if $q$ is false at the closest $p$-worlds. So von Fintel’s and Gilles’s semantics can be weaker than Lewis’s here.

(Why care about the truth-value of non-uttered conditionals? There are several reasons to do so, but a salient one in the present discussion is that conditionals are used in various analyses of knowledge. This may have the consequence that a knowledge ascription uttered in a given context depends on the truth-value of some non-uttered conditional.)

### A.4 Hybrid semantics

Williamson (2000, 149) sketches a different non-standard semantics for subjunctive conditionals, which he calls “Nozickian”. I call it “hybrid” for reasons that should shortly be clear.

The semantics uses the Lewis-style ordering in terms of comparative closeness. But it also uses a binary close/distant relation between worlds. To every world is associated a sphere $S^C_w$ of close worlds. We write $wCw'$ when $w'$ is close to $w$. Closeness is akin to an accessibility relation. Plausibly, it is coordinated with the comparative closeness relation so that any world closer than a close world is also close: $\forall w, w', w'' ((wCw'' \land w' \leq_w w'') \rightarrow wCw')$. Thus binary closeness may be thought of as a fixed threshold on the Lewis ordering.

On the hybrid semantics, if the antecedent is false, we have a Lewis-style variable truth condition:

**Hybrid semantics (1)** If $p$ is false at $w$, $p \not\rightarrow q$ is true at $w$ iff $p \rightarrow q$ holds throughout some $p$-permitting sphere in $S_w$, or there is no $p$-permitting sphere in $S_w$.

---

12. Williamson discusses the semantics on behalf of Nozick and Sosa, who need non-trivially true conditionals with true antecedents for their adherence and safety condition on knowledge, respectively. He uses the fixed-sphere apparatus for his own safety condition, but as a semantics for “$p$ could easily have happened” claims rather than for subjunctive conditionals.
If the antecedent is true, we have a fixed but non-contextualist semantics:

**Hybrid semantics (2)** If $p$ is true at $w$, $p \rightarrow q$ is true at $w$ iff $p \rightarrow q$ holds throughout $S^C_w$, where $S^C_w$ is the sphere of worlds close to $w$.

The second clause involves a fixed sphere for all true antecedents. It is thus akin to the contextualist fixed semantics, except that the sphere does not even depend on context. (Of course we can also consider a contextualist hybrid semantics.)

The semantics is hybrid because it is variable for true antecedents and fixed for false ones, but also because it is existential for true antecedents and singular or universal for false ones. Instead of a having a unique close-worlds sphere $S_w$, we could equivalently formulate the truth-condition for true antecedents in terms of a system of close spheres that would be a subset of the full system of spheres, namely $S^C_w = \{ S \in S_w : S \subseteq S^C_w \}$. But to get the intended truth-condition, we have to require $p \rightarrow q$ to hold throughout all close spheres.

When the antecedent is false, we consider the sphere of closest worlds up to the closest $p$-worlds and no further. That may be more than the sphere of close worlds, if the closest $p$-worlds are not close. But that may also be less, for some $p$-worlds may be close without being the closest ones. When the antecedent is true, we consider the close worlds and only them. That will typically be more that the closest $p$-worlds, since when $p$ holds in $w$ the closest $p$-worlds are only the ones as close to $w$ as $w$ itself.

The logic is restricted to principles valid both for the fixed and the variable semantics. It is not the case that conditionals with true antecedents get the logic of the fixed strict conditional. For them, transitivity does hold: if $p$ and $q$ are true, $p \rightarrow q$ and $q \rightarrow r$ entail $p \rightarrow r$. But contraposition is limited: if $p$ is true, then $p \rightarrow q$ contraposes only if $q$ is false. If $q$ is false, $\neg q \rightarrow \neg p$ has a true antecedent, and so relies on the sphere of close worlds as well; but if $q$ is true, the truth of $\neg q \rightarrow \neg p$ depends on where the closest $\neg q$ worlds are. And of course transitivity and contraposition fail in general, since they do not hold for conditionals with false antecedents.
A.5 Nozick’s tracking

A.5.1 Simple version

Recall Nozick’s (1981, 172–7) simple sensitivity and adherence conditions:

**Sensitivity**  If \( p \) were not true, \( S \) would not believe that \( p \).

**Adherence**  If \( p \) were true, \( S \) would believe that \( p \).

Nozick’s weakened spread semantics is:

**Nozick’s Weakened Spread semantics**  \( p \rightarrow q \) is true at \( w \) iff \( p \rightarrow q \) holds at all spheres in \( S_w \) whose \( p \) worlds are all in the initial part of the closest \( p \) spread.

Equivalently, \( p \rightarrow q \) must hold in spheres that include at most the initial part of the closest \( p \) spread. Nozick’s semantics is quantified, but it is easily turned into a singular semantics: let the *initial \( p \)-spread sphere* be the intersection of all spheres including at most the initial part of the closest \( p \) spread. We have:

**Nozick’s Weakened Spread semantics**  \( p \rightarrow q \) is true at \( w \) if \( p \rightarrow q \) holds throughout the initial \( p \)-spread sphere of \( w \).

Given this semantics, Nozick’s tracking conditions can be formulated as simple modal requirements:

**Nozickian sensitivity (NS)**  For any \( \alpha \), \( NS_{\alpha} \) iff \( B_{\alpha} \) and \( \neg I_{\beta} \) for any \( \beta \) such that \( \alpha R_{\beta} \), where:

\[
\alpha R_{\beta} \text{ iff } p_{\beta} = p_{\alpha}, \ S_{\beta} = S_{\alpha}, \text{ and } w_{\beta} \text{ is in the initial } \neg p_{\alpha}\text{-spread sphere of } w_{\alpha}.
\]

**Nozickian adherence (NA)**  For any \( \alpha \), \( NA_{\alpha} \) iff \( B_{\alpha} \) and \( \neg E_{\beta} \) for any \( \beta \) such that \( \alpha R'_{\beta} \), where:

\[
\alpha R'_{\beta} \text{ iff } p_{\beta} = p_{\alpha}, \ S_{\beta} = S_{\alpha}, \text{ and } w_{\beta} \text{ is in the initial } p_{\alpha}\text{-spread sphere of } w_{\alpha}.
\]

As desired, \( R \) and \( R' \) are reflexive, since all spheres of cases around \( \alpha \) include \( \alpha \). The two conditions cannot be unified, since the spheres
containing the initial \(p_\alpha\) spread and the initial \(\neg p_\alpha\) spread will typically not coincide.

Nozickian sensitivity is slightly stronger than sensitivity as usually understood (SE). Nozickian adherence is typically stronger than simply requiring true belief, and weaker than a Strict Spread requirement, but cannot be compared to Quasi-Nozickian or Stretch semantics.

Not all quantified semantics for conditionals can be reduced to singular ones. With essentially quantified ones, Nozick’s tracking conditions into quantified modal requirements. Instead of relying on a single relation \(R\) between cases, and thus a single “sphere” of relevant alternatives, a quantified requirement quantifies over a system of such spheres:

**Quantified modal requirements schema**

*Universal.* For any \(\alpha\), \(QMR_\alpha\) iff \(B\alpha\) and \(\neg E\alpha\) holds throughout all spheres in \(S_\alpha\) satisfying condition \(F(\alpha)\).

*Existential.* For any \(\alpha\), \(QMR_\alpha\) iff \(B\alpha\) and \(\neg E\alpha\) holds throughout some sphere in \(S_\alpha\) satisfying condition \(F(\alpha)\).

Recall the Quasi-Nozickian semantics:

**Quasi-Nozickian semantics** \(p \square \rightarrow q\) is true at \(w\) iff: either (a) \(p \rightarrow q\) holds throughout some sphere in \(S_w\) that permits both \(p\) and \(\neg p\), or (b) \(p \rightarrow q\) holds at all worlds in \(\bigcap S_w\).

To state the Quasi-Nozickian version of sensitivity and adherence, we build a system of spheres of cases on the basis of the system of spheres of worlds. Let \(S_\alpha\) be the set of cases spheres \(S\) such that for some worlds sphere \(S' \in S_{w_\alpha}\), \(S'\) permits both \(p_\alpha\) and \(\neg p_\alpha\), and for any \(\beta, \beta' \in S\) iff \(p_\beta = p_\alpha\), \(S_\alpha = S_\beta\) and \(w_\beta \in S'\). For any worlds sphere \(S'\) containing both \(p_\alpha\) and \(\neg p_\alpha\) worlds, the cases sphere \(S\) captures all cases in \(S'\) that involve the same subject and proposition as \(\alpha\). If no sphere in \(S_{w_\alpha}\) permits both \(p_\alpha\) and \(\neg p_\alpha\), let \(S_\alpha\) contain simply the cases sphere \(S\) such that for any \(\beta, \beta' \in S\) iff \(p_\beta = p_\alpha\), \(S_\alpha = S_\beta\) and \(w_\beta \in \bigcap S_{w_\alpha}\). This allows a unified statement of sensitivity and adherence:

**Quasi-Nozickian tracking (QNT)** For any \(\alpha\), \(QNT_\alpha\) iff \(B\alpha\) and there is some sphere in \(S \in S_\alpha\) such that \(B\beta \leftrightarrow T\beta\) for any \(\beta \in S\).
That is, ignorance and error must be avoided over a sphere of cases permitting both \( p_a \) and \( \neg p_a \) cases, where ignorance and error consists in \( S_a \)'s mistakenly believing \( p_a \) or failing to believe truly \( p_a \). If \( p_a \) is a necessary truth, we have a unique sphere \( S_a \) containing all cases in \( S_w \) that involve the same subject and proposition: sensitivity is trivially satisfied, for \( p_a \) is a necessary truth and no error can occur at any case, but adherence requires avoidance of ignorance at all such cases. This latter condition is likely not to be satisfied, which illustrates the sceptical consequences of adherence on some variable semantics.

A.5.2 Nozick’s method-relative tracking

Recall Nozick’s (1981, 179) method-relative tracking conditions. Assuming that \( S \) believes that \( p \) via method \( m \):

Nozick’s sensitivity with methods If \( p \) was not true and \( S \) were to use \( m \) to arrive at a belief whether (or not) \( p \), then \( S \) would not believe, via \( m \), that \( p \).

Nozick’s adherence with methods If \( p \) was true and \( S \) were to use \( m \) to arrive at a belief whether (or not) \( p \), then \( S \) would not believe, via \( m \), that \( p \).

We abbreviate:

\[ B_{m\alpha} \text{ in } \alpha, \text{ } S_{\alpha} \text{ believes that } p_a \text{ on the basis of } m. \]

\[ E_{m\alpha} \text{ in } \alpha, \text{ } S_{\alpha} \text{ is mistaken on the basis of } m: B_{m\alpha} \land \neg T_{\alpha}. \]

\[ I_{m\alpha} \text{ in } \alpha, \text{ } S_{\alpha} \text{ is ignorant on the basis of } m: T_{\alpha} \land \neg B_{m\alpha}. \]

\( S_a m?p_a \) \( S_a \) uses \( m \) to arrive at a belief whether \( p_a \).

On Nozick’s weakened spread semantics, we get parametric modal requirements:

Nozickian M-sensitivity (NMS) For any \( \alpha \), \( NMS_\alpha \) iff there is an \( m \) such that \( B_{m\alpha} \) and \( \neg E_{m\beta} \) for any \( \beta \) such that \( \alpha R_{m\beta} \), where:

\( \alpha R_{m\beta} \) iff \( p_\beta = p_a, \text{ } S_\beta = S_\alpha, \text{ and } w_\beta \text{ is in initial } \neg p_a \land S_a m?p_a\text{-spread sphere of } w_\alpha. \)
A.5. Nozick’s tracking

Nozickian M-adherence (NMA) For any \( \alpha \), \( N\alpha \) iff \( B\alpha \) and \( \neg I\beta \) for any \( \beta \) such that \( \alpha R'_m\beta \), where:
\[
\alpha R'_m\beta \text{ iff } p_\beta = p_\alpha, S_\beta = S_\alpha, \text{ and } w_\beta \text{ is in the initial } p_\alpha \land S_\alpha m?p_\alpha\text{-spread sphere of } w_\alpha.
\]

\( R_m \) and \( R'_m \) are reflexive, as desired. For any \( \alpha \), \( (B\alpha \land \neg E\alpha) \rightarrow T\alpha \), so sensitivity requires true belief, again as desired.

On the revision suggested by Luper-Foy (1984, 28–9) and Williamson (2000, 154), the antecedent of the Sensitivity conditional should not refer to methods. In Nozickian semantics, revised method-relative Sensitivity is formulated:

Revised-Nozickian M-sensitivity (RNMS) For any \( \alpha \), \( N\alpha \) iff there is an \( m \) such that \( Bm\alpha \) and \( \neg E\alpha \) for any \( \beta \) such that \( \alpha R_m\beta \), where:
\[
\alpha R_m\beta \text{ iff } p_\beta = p_\alpha, S_\beta = S_\alpha, \text{ and } w_\beta \text{ is in initial } \neg p_\alpha\text{-spread sphere of } w_\alpha.
\]

As we pointed out, an analogous revision of method-relative Adherence is ill-advised (sec. 3.5.3). So Revised Nozickian method-relative tracking would simply substitute the revised sensitivity clause for the original one above.

To illustrate a quantified version of Nozick’s method-relative tracking, we use the Quasi-Nozickian semantics again. The semantics requires spheres that permit both truth and falsity of the antecedent. Method-relative sensitivity requires a sphere that permits both \( \neg p_\alpha \land S_\alpha m?p_\alpha \) and the negation of it, namely \( p_\alpha \lor \neg S_\alpha m?p_\alpha \). Method-relative adherence requires a sphere that permits both \( p_\alpha \land S_\alpha m?p_\alpha \) and the negation of it, \( \neg p_\alpha \lor \neg S_\alpha m?p_\alpha \). Now when \( \alpha \) is not a case of belief, the grounding condition — that \( Bm\alpha \) holds for some \( m \) — will automatically fail, and when \( p_\alpha \) is false, sensitivity will automatically fail. \(^{13}\) So we can focus on the interesting case in which \( p_\alpha \) is true and believed in \( \alpha \). For sensitivity,

\(^{13}\) If \( p_\alpha \) is false, for any \( m \) such that \( Bm\alpha \), the actual world permits \( \neg p_\alpha \land S_\alpha m?p_\alpha \). Our spheres of cases will contain either various spheres that also permit its negation, or simply all cases involving the same subject and proposition. However this turns out to be, all spheres contain the actual world, which is a case of error based on \( m \), so sensitivity fails.
any sphere will allow \( p_\alpha \lor \neg S_\alpha m ? p_\alpha \), since \( p_\alpha \) holds in \( \alpha \). So the condition boils down to finding spheres that allow \( \neg p_\alpha \land S_\alpha m ? p_\alpha \): roughly, we go up to the first worlds in which \( p_\alpha \) is false but \( S_\alpha \) still uses \( m \) to settle whether \( p_\alpha \). For adherence, any sphere will allow \( p_\alpha \land S_\alpha m ? p_\alpha \) for any \( m \) such that \( Bm_\alpha \), since \( p_\alpha \) is true in \( \alpha \). So the condition boils down to finding spheres that allow \( \neg p_\alpha \lor S_\alpha m ? p_\alpha \): roughly, we go up to the first \( \neg p_\alpha \) cases, or to the first cases in which \( S_\alpha \) does not use \( m \) to settle whether \( p_\alpha \), depending on which are closest. We have to define two systems of spheres of cases:

**Spheres for Quasi-Nozickian M-tracking**

*Sensitivity.* \( S_{m,\alpha} \) is the set of cases spheres \( S \) such that for some world sphere \( S' \in S_w \), \( S' \) permits both \( \neg p_\alpha \land S_\alpha m ? p_\alpha \) and \( p_\alpha \lor \neg S_\alpha m ? p_\alpha \), and for any \( \beta \), \( \beta \in S \) iff \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \) and \( w_\beta \in S' \). If no sphere in \( S_w \) permits both \( \neg p_\alpha \land S_\alpha m ? p_\alpha \) and \( p_\alpha \lor \neg S_\alpha m ? p_\alpha \), let \( S_{m,\alpha} \) contain only the sphere \( S \) such that for any \( \beta \), \( \beta \in S \) iff \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \) and \( w_\beta \in \bigcap S_{w_\alpha} \).

*Adherence.* \( S'_{m,\alpha} \) is the set of cases spheres \( S \) such that for some world sphere \( S' \in S_{w_\alpha} \), \( S' \) permits both \( p_\alpha \land S_\alpha m ? p_\alpha \) and \( \neg p_\alpha \lor \neg S_\alpha m ? p_\alpha \), and for any \( \beta \), \( \beta \in S \) iff \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \) and \( w_\beta \in S' \). If no sphere in \( S_w \) permits both \( p_\alpha \land S_\alpha m ? p_\alpha \) and \( \neg p_\alpha \lor \neg S_\alpha m ? p_\alpha \), let \( S_{m,\alpha} \) contain only the sphere \( S \) such that for any \( \beta \), \( \beta \in S \) iff \( p_\beta = p_\alpha \), \( S_\beta = S_\alpha \) and \( w_\beta \in \bigcap S_{w_\alpha} \).

**Quasi-Nozickian tracking (QNMT)**

For any \( \alpha \), \( QNMT \alpha \) iff there is some \( m \) such that \( Bm_\alpha \) and there is some sphere in \( S \in S_\alpha \) such that \( \neg E \beta \) for any \( \beta \in S \) and some sphere \( S' \in S'_{m,\alpha} \) such that \( \neg I \beta \) for any \( \beta \in S \).
Appendix B

On the “warrant entails truth” argument

My generalisation of Gettier cases against fallibilist requirements is reminiscent of Merrick’s argument that “warrant entails truth” (Merricks, 1995, 841). Let me quickly point out why I find Merrick’s formulation unsatisfactory.

Following Plantinga, Merricks defines warrant as “that which makes the difference between knowledge and mere true belief” (Plantinga, 1993, 3). Merricks (1995, 842) calls infallibilism the view that “warrant entails truth”, and fallibilism the view that it does not (842n3). The claim is meant to rule out a number of accounts of knowledge (Merricks, 1995, 854n27), and has consequently generated some debate.¹

The debate is prima facie surprising, since the claim that warrant entails truth does not rule out any account of knowledge. Consider the justified true belief analysis. On the Chisholmian view, justification does not entail truth. So it seems that the traditional justified true belief analysis is ruled out by the claim. But it is not. For we can propose that knowledge is warranted true belief, where a belief is warranted iff it is justified and true. This “hidden” justified true belief analysis is equivalent to the original one, but sports a notion of warrant on which warrant entails truth.

A second issue is that Merricks does not actually defend the claim that warrant entails truth, but that warranted belief entails truth. His

¹ Ryan (1996); Howard-Snyder et al. (2003); Coffman (2008) reject it, Merricks (1997) defends it.
argument is a *reductio* of the hypothesis that there is a warranted but false belief (Merricks, 1995). It is not a *reductio* of the hypothesis of a warranted falsehood. Many would think the difference harmless, because they assume that where there is warrant, there is belief: one has to be warranted in *something*. But that is already leaving the strictly logical notion of warrant we started with. To illustrate: Nozick’s sensitivity (if \( p \) was false, \( S \) would not believe \( p \)) can be true in a case in which \( S \) does not believe \( p \). By itself, it does not entail truth, but in conjunction with belief, it does. In subsequent discussion, many have (understandably) taken Merricks to have argued that *warrant*, not warranted belief, entailed truth (Huemer, 2005, 171, 176).

Our first issue, the trivialisation of the debate, reveals a problem with the notion of warrant pointed out by Huemer (2005, 172–3): the definition of warrant does not guarantee uniqueness. In our terminology of cases (3.1.1) we have:

**Warrant**  \( W \) is a warrant condition iff for any case \( \alpha \), \( K\alpha \) iff \( B\alpha \land T\alpha \land Wa \).

It follows from the definition that knowledge is itself a warrant condition. But more damagingly, many mutually non-equivalent conditions are warrant conditions. Take any warrant condition \( W \). Define \( W’ \) and \( W'' \) as:

\[
W’ \text{ is such that for any case } \alpha, W’\alpha \text{ iff } Wa \lor \neg(B\alpha \land Ta). \tag{3}
\]

\[
W'' \text{ is such that for any case } \alpha, W''\alpha \text{ iff } Wa \land (B\alpha \land Ta).
\]

\( W’ \) and \( W'' \) are warrant conditions if \( W \) is. \( W \) entails \( W’ \) and is entailed by \( W'' \). But if there are cases \( \beta \) such that \( \neg W\beta \land \neg B\beta \), we have \( W’\beta \) but not \( W\beta \), so \( W’ \) is weaker than \( \beta \). And if there are cases such that \( W\alpha \land \neg B\alpha \), \( W'' \) is stronger than \( W \). More generally, if \( W \) is a warrant condition:

any condition \( W’ \) such that \( W’ = W \lor C \) where \( C \) entails \( \neg (B \land T) \) is also a (typically weaker) warrant condition, and

2. Suppose \( Ka \). Since knowledge entails belief and truth, we have \( B\alpha \land Ta \land Ka \). Now suppose \( B\alpha \land Ta \land Ka \). Trivially, we have \( Ka \). So for any \( \alpha \), \( Ka \equiv B\alpha \land Ta \land Ka \).


4. By propositional logic, both \( B\alpha \land Ta \land Wa \lor \neg(B\alpha \land Ta) \) and \( B\alpha \land Ta \land Wa \land (B\alpha \land Ta) \) are equivalent to \( B\alpha \land Ta \land Wa \).
any condition \( W'' = W \land C \) where \( (B \land T) \) entails \( C \) is a (typically stronger) warrant condition. 5

Given the unrestricted definition of warrant, the question whether warrant entails truth loses its point (Huemer, 2005, 176). Huemer shows it for the claim that warrant entails truth, but we can equally show it for Merrick’s actual claim that warranted belief entails truth. If the question is understood as saying that on all warrant conditions, warranted belief entails truth, it is trivially false. Suppose \( W \) is a warrant condition such that \( B \land W \) entails truth. Then \( W' = W \lor \neg T \) is a warrant condition such that \( B \land W' \) does not entail truth. If it is understood as saying that some warrant condition entails truth, it is trivially true. Suppose \( W \) is a warrant condition such that \( B \land W \) does not entail truth. Then \( W'' = W \land T \) is a warrant condition such that \( B \land W'' \) entails truth — just as our justified and true condition was.

Bailey (forth., 6–8) provides guidelines to narrow down the notion of warrant. He defines warrant as the **logically weakest epistemic good** that satisfies the previous definition:

**Real Warrant** \( W \) is a real warrant condition iff (1) \( W \) is a warrant condition,

(2) \( W \) is an epistemic good, (3) there is no warrant condition \( W' \) such that \( W \) entails \( W' \) but not conversely and \( W' \) is an epistemic good.

Suppose \( W \) is a real warrant condition. Then the stronger conditions \( W \land T \) or \( W \land (B \lor T) \) are not real warrant conditions, for there is a logically weaker condition that is a warrant condition and an epistemic good, \( W \). On the other hand, weaker conditions such as \( W \lor \neg T \) or \( W \lor \neg (B \land T) \) are not real warrant conditions because they are not proper epistemic goods. For they can be satisfied by an irrational false belief, which is not an epistemic good. Generalising: any condition \( W \lor C \) such that \( C \) entails \( \neg (B \land T) \) is equivalent to \( W \) or not; if not, there are cases \( \beta \) where it is satisfied by \( C \) alone: \( C\beta \land \neg W\beta \). But since \( C \) entails the absence of a true belief, it cannot be good. Bailey assumes here (1) that an epistemic good is a

---

5. \( W' = W \lor C \) abbreviates: for any \( \alpha \), \( W'\alpha \equiv W\alpha \lor C\alpha \). “\( C \) entails \( \neg (B \land T) \)” abbreviates: for any \( \alpha \), if \( C\alpha \) then \( \neg (B\alpha \land T\alpha) \).
condition such that any case satisfying it is epistemically good in some respect; (2) the absence of a true belief is not epistemically good. The weak points of Bailey’s reformulation are the notion of epistemic good, which is unclear, and the idea that the absence of a true belief is not epistemically good, which is not obvious.

Let us look at warrant from the point of view of methods infallibilism.

Write:
\[ K_{\alpha} \text{ In } \alpha, S_{\alpha} \text{ knows } p_{\alpha} \]
\[ B_{m\alpha} \text{ In } \alpha, S_{\alpha} \text{ believes } p_{\alpha} \text{ on the basis of } m \]
\[ T_{\alpha} \text{ In } \alpha, p_{\alpha} \text{ is true} \]
\[ E_{m\alpha} \text{ } B_{m\alpha} \land \neg T_{\alpha}. \]
\[ a R \beta \text{ case } \beta \text{ is accessible from } \alpha \text{ — i.e., it is a relevant error possibility. } R \text{ is reflexive.} \]

Methods infallibilism, in its “radical” form on which method infallibility is necessary and sufficient for knowledge, is the claim that:

**Radical methods infallibilism** For any \( \alpha \), \( K_{\alpha} \iff \exists m (B_{m\alpha} \land \forall \beta (a R \beta \rightarrow \neg E_{m\beta})). \)

We cannot extract independent truth and belief conditions from the requirement. (Note in particular that the belief operator is under the scope of the existential quantifier.) The only way to fit the view in the Warrant pattern is:

For any \( \alpha \), \( K_{\alpha} \iff B_{\alpha} \land T_{\alpha} \land \exists m (B_{m\alpha} \land \forall \beta (a R \beta \rightarrow \neg E_{m\beta})), \)

where truth and belief are redundant, given the reflexivity of \( R \).

We can extract the belief condition if we focus on knowledge on a basis.

Write:
\[ K_{m\alpha} \text{ In } \alpha, S_{\alpha} \text{ knows } p_{\alpha} \text{ on the basis of } m. \]

We have:

For any \( \alpha, m \), \( K_{m\alpha} \iff B_{m\alpha} \land \forall \beta (a R \beta \rightarrow \neg E_{m\beta}). \)

Or equivalently, given the reflexivity of \( R \):

For any \( \alpha, m \), \( K_{m\alpha} \iff B_{m\alpha} \land T_{\alpha} \land \forall \beta (a R \beta \rightarrow \neg E_{m\beta}). \)
On the “warrant entails truth” argument

The closest analogue of Warrant in that requirement is the infallibility of $m$, $\forall \beta (aR\beta \rightarrow \neg E_m\beta)$. It is not a condition that entails truth by itself. But it does so in conjunction with belief. So it satisfies Merrick’s idea that warranted belief entails truth, but not the idea that warrant entails truth.

I am not sure whether the infallibility of $m$ is an “epistemic good” in all cases in which it obtains. For it obtains in cases in which the subject does not “have” the method in question. Let there be two subjects $a$ and $b$. In $\alpha$, $a$ believes $p_\alpha$ on the basis of sight, and sight is infallible. In $\beta$, $b$ believes $p_\beta$ on the basis of hearing, and hearing is fallible. Suppose each case is a relevant error possibility for each other: then in $\beta$ it is true that sight is infallible, but that does no good to the subject of $\beta$, $b$.

The methods infallibilist version of warrant also shows what is wrong with Blome-Tillmann’s (2007, 216–8) argument that warrant entails belief. Blome-Tillmann’s crucial premise is:

(B.1) If $S$ has warrant to believe $p$ without actually believing $p$, then there is a close possibility in which $p$ is true, $S$ equally has warrant to believe it, but $S$ comes to believe it by some irrational method such as reading tea leaves.

The problem with the argument lies with the undefined notion of “being warranted to believe $p$”. There is no equivalent of such notion in the methods account. We have a notion of being warranted to believe on a certain method:

for any $\alpha$, $\alpha$ is a case of being warranted to believe on the basis of $m$ iff:

$\forall \beta (aR\beta \rightarrow \neg E_m\beta)$.

Being warranted to believe on the basis of $m$ in $\alpha$ is another way of saying that $m$ is infallible in $\alpha$. Being warranted to believe on the basis of $m$ does not entail belief. In intuitive terms, the condition is typically satisfied when counterfactuals of the following form hold: if $S$ were to believe something on the basis of $m$, that belief would be true.

Suppose $\alpha$ is a case of being warranted to believe on the basis of $m$ without belief:

$\neg B\alpha \land \forall \beta (aR\beta \rightarrow \neg E_m\beta)$.
Suppose that there is a close case $\gamma$ in which the subject of $\alpha$ comes to believe the proposition of $\alpha$ on the basis of reading tea leaves. Write $n$ for the tea leaves method. We have:

$$Bn\gamma \land \forall \beta (\gamma R \beta \rightarrow \neg E_m \gamma).$$

This is a case of one’s believing on the basis of $n$ and $m$ being infallible. Since the belief is not based on the infallible method, it does not follow that the methods infallibilist conditions on knowledge are satisfied. Blome-Tillman’s premise (B.1) fails.

Relatedly, Blome-Tillmann (2007, 216n) states that “most familiar substantive accounts [such as] causal, reliabilist, counterfactual and safety-based accounts directly entail the thesis that warrant entails belief.” This is mistaken, as our formalisations of the accounts in ch. 3 show. For instance, the variants of safety and sensitivity that do not appeal to method yield accounts of knowledge of the form:

For any $\alpha$, $K\alpha$ iff $B\alpha \land T\alpha \land \forall \beta (\alpha R \beta \rightarrow \neg E\beta)$.

where $E\beta$ iff $B\beta \land \neg T\beta$ and $R$ is reflexive, so that the truth condition is redundant.

Warrant can be identified with the requirement that holds in any $\alpha$ iff $\forall \beta (\alpha R \beta \rightarrow \neg E\beta)$, which does not entail belief. In intuitive terms, such requirements are roughly stating that: if $S$ were to form a belief that $p$, it would not be a false belief. On that account, warrant does not entail truth, though warranted belief does.

Modal requirements in general state that one is in a kind of situation where certain systematic modal relations obtain between belief and truth. The existence of these relations entails neither actual belief nor actual truth. But it entails that if there is a belief (of the right type), it must be true.

The confusions over whether warrant or warranted belief entail truth and over the notion of being warranted to believe illustrate how easy it is to assume more about the notion of warrant (in particular, to equate it with being justified) than its explicit definition allows.
Appendix C

On Methods Models

C.1 Algebra for methods

\((M, +, \circ, 0, 1)\) is an algebraic structure over the set of methods, where
0 and 1 are defined as follows:

Definition C.1. The empty method, noted 0, is the method such that
0\((w, \pi) = \emptyset\) for all \(w, \pi\).

The identity method, noted 1, is the method such that 1\((w, \pi) = \pi\) for
all \(w, \pi\).

Here are the main properties of the algebra.

Theorem C.1. Method union is idempotent, commutative and associative.

For any \(m, n, r \in M\): \(m + m = m\), \(m + n = n + m\), and \((m + n) + r = m + (n + r)\).

Proof. From the corresponding properties of set union and Definition
5.2. \(\square\)

Remark C.1. The empty method 0 is uniquely characterised as the method
such that 0 + \(n = n\) for any \(n\).

Theorem C.2. Method composition is associative but not idempotent nor com-
mutative.

For any \(m, n, r \in M\): \(m \circ (n \circ r) = (m \circ n) \circ r\). But \(m \circ m = m\) and
\(m \circ n = n \circ m\) are not valid.
Proof. Associativity: from the associativity of function composition and Definition 5.2.

Counterexample to idempotence: for any proposition \( p \in P \), write \( \neg p \) the negation of \( p \). Let \( m \) be such that for any \( w, \pi \), \( m(w, \pi) = \{ \neg p : p \in \pi \} \). At any \( w \) we have: \( m(w, \{p\}) = \{\neg p\} \neq (m \circ m)(w, \{p\}) = \{\neg\neg p\}. \)

Counterexample to commutativity: consider \( m \) defined as above, and \( n \) such that at any \( w \), \( n(w, \pi) = \{ p \land q : p, q \in \pi \} \) where \( p \land q \) denotes the conjunction of any propositions \( p \) and \( q \). Assuming \( p \neq q \), we have \( \neg(p \land q) \in (m \circ n)(w, \{p, q\}) \) but \( \neg(p \land q) \notin (n \circ m)(w, \{p, q\}) \) at any \( w \), so \( m \circ n \neq n \circ m \). \( \square \)

Remark C.2. The identity method \( 1 \in M \) such that \( 1(w, \pi) = \pi \) for any \( w, \pi \) is uniquely characterised as the method such that \( 1 \circ n = n \circ 1 = n \) for any \( n \in M \).

Remark C.3. Purely non-inferential methods are methods which are insensitive to what premises they are given. They can thus be characterised as the set of methods \( m \) such that \( m \circ n = m \) for any \( n \in M \).

For any \( n \in M, w, \pi: 0 \circ n = 0 \) (the empty method is purely non-inferential) and \( (n \circ 0)(w, \pi) = n(w, \emptyset) \) (combining a method \( n \) with the empty one amounts to applying \( n \) without premise).

Theorem C.3. Method union does not distribute over method combination.

\[ m + (n \circ r) = (m + n) \circ (m + r) \text{ is not valid.} \]

Proof. Take \( r = 1; \) the claim reduces to \( m + n = (m + n) \circ (m + 1) \), which is guaranteed only if \( m + n \) is non-inferential. \( \square \)

Theorem C.4. Composition distributes right-to-left over union, but not left-to-right.

For any \( m, n, r \in M: (m + n) \circ r = (m \circ r) + (n \circ r) \). By contrast, \( m \circ (n + r) = (m \circ n) + (m \circ r) \) is not valid.

1. The counterexamples given in this section assume a few uncontroversial facts about propositions, such as: the negation of a proposition is a proposition and at least some negation of a proposition is distinct from its own negation. These will hold however propositions are fleshed out.
C.2. Comparison with neighbourhood models

Proof. Left-to-right distribution. For any \( w, \pi, (m + n)(w, \pi) = m(w, \pi) \cup n(w, \pi) \) (Definition 5.2). Consider \( \pi = r(w, \pi') \) for any given \( \pi' \).

Counterexample to right-to-left distribution. Write \( p \lor q \) for the disjunction of any propositions \( p \) and \( q \). Let \( m \) be such that \( m(w, \pi) = \{ p \lor q : p, q \in \pi \} \). Consider \( w, n, r \) such that \( n(w, \emptyset) = \{ p \} \) and \( r(w, \emptyset) = \{ q \} \). Assuming \( p \neq q \), we have \( m \circ (n + r)(w, \emptyset) = \{ p \lor q, p, q \} \neq (m \circ n) + (m \circ r)(w, \emptyset) = \{ p, q \} \). □

Composition distributes right-to-left but not left-to-right because methods algebra represent information flow or informational dependencies. Composing \( m \) with \( n + r \) means that \( m \) can use the outputs of \( n \) and \( r \) together; this is not the same as applying \( m \) to the outputs of \( r \) and and those of \( n \) separately. So typically, \( m \circ (n + r)(w, \emptyset) \neq (m \circ n) + (m \circ r) \). By contrast, pooling together the \( m \)- and \( n \)-inferences and applying them to a single output is the same as applying \( m \) and \( n \) separately to that output, so \( (m + n) \circ r = (m \circ r) + (m \circ r) \).

To sum up, we have an algebra \( \langle M, +, \circ, 0, 1 \rangle \) with two distinguished elements, the empty method (identity element for \( + \)) and the identity method (identity element for \( \circ \)). \( + \) is associative, commutative and idempotent, \( \circ \) is associative. \( \circ \) distributes right-to-left over \( + \) but not left-to-right.

C.2 Comparison with neighbourhood models

Method models amount to building a neighbourhood model with two modalities out of the agent’s basic methods and a background alethic modality. They are richer than simple neighbourhood models because they give insight on a structure of methods (built out of composition and union) which is, so to speak, the scaffolding with which the neighbourhood functions for knowledge and belief are built. That is why our models are more explanatory, as we will see.

A neighbourhood frame \( \mathcal{F} \) is a pair \( \langle W, N \rangle \) where \( W \) is a set of worlds
and \( N \subseteq W \times \mathcal{P}(W) \) a function from worlds to sets of propositions.\(^2\)

**Definition C.2.** Let \( \mathcal{L}_\nu \) be the set of formulas given by:

\[
\phi ::= p | \top | \neg \phi | \psi | \nu \phi
\]

where \( P = \{p, q, \ldots\} \) is a set of propositional constants.

Let \( M = \langle F, V \rangle \) be a model where \( V : P \rightarrow \mathcal{P} \) is a valuation function. We define \( \llbracket \cdot \rrbracket_M \):

\[
\llbracket p \rrbracket_M = V(p), \\
\llbracket \neg \phi \rrbracket_M = W \setminus \llbracket \neg \phi \rrbracket_M \\
\llbracket \phi \lor \psi \rrbracket_M = \llbracket \phi \rrbracket_M \cup \llbracket \psi \rrbracket_M \\
\llbracket \nabla \phi \rrbracket_M = \{w : \llbracket \phi \rrbracket_M \in N(w)\}.
\]

(And as usual for other logical connectives)

**Truth.** \( \models_w^M \phi \) iff \( w \in \llbracket \phi \rrbracket_M \).

**Validity.** \( \models^M \phi \) iff for any world \( w, \models_w^M \phi \).

In methods models, methods are functions from worlds to inference transitions functions, that is functions from sets of premises to sets of conclusions. For a given set of premises \( \pi \) and a given method \( m \), the function \( w \mapsto m(w, \pi) \) is a function from worlds to sets of conclusions. If propositions are sets of possible worlds, this is a neighbourhood function. In particular, the unconditional output function \( w \mapsto m(w) \) of a method \( m \) is a neighbourhood function.\(^3\) And so are the functions \( B(m, w), K(m, w), B(w) \) and \( K(w) \) that we build out of them.

Let \( \mathcal{L}_B \) be the restriction of our methods language \( \mathcal{L} \) to formulas that contain at most the B operator (as well as propositional connectives) and \( \mathcal{L}_K \) its restriction to formulas that contain at most K operator. We can establish two useful equivalence results:\(^4\)

**Theorem C.5.** For each method frame \( \mathcal{F} \), there is a neighbourhood frame \( F \) that is pointwise equivalent over \( \mathcal{L}_B \), and conversely.

---


3. Recall that \( m(w) \) abbreviates \( m(w, \emptyset) \).

4. Thanks to Johannes Stern here.
C.2. Comparison with neighbourhood models

Proof. Let ≰ = < W, M^R, R > be any methods frame. Define the neighbourhood frame F = < W, N > such that for any w, N(w) = B(w) and for any model M in F, [Bφ]^M = {w : [φ]^M ∈ N(w)} . We prove that F is pointwise equivalent to ≰ over L_B by induction on the complexity of φ.

The interesting case is:

\[ \models^w_B φ \text{ iff } [φ]^M ∈ N(w) (\text{semantics}) \]
\[ \text{iff } [φ]^M ∈ B(w) (\text{definition of N}) \]
\[ \text{iff } [φ]^M ∈ N(w) (\text{inductive hypothesis}) \]
\[ \text{iff } [φ]^M ∈ B(w) (\text{definition of union}) \]

Conversely let F = < W, N > be any neighbourhood frame for B. Define the methods frame ≰ = < W, M^R, R > such that M^R = {m} where m is such that for any \( w, \pi : m(w, \pi) = N(w) \) and R some reflexive accessibility relation. We first prove that M = \{m\}: \( m + m = m \) and \( m \circ m = m \) (the first holds for any method by the definition of union (Definition 5.2), the second holds because the definition of m entails that \( m(w, m(w, \pi)) = N(w) = m(w, \pi) \) for any \( \pi \); and since \( M^R = \{m\} \), \( M = M^{R+} = \{m\} \) (Definition 5.6). We then prove that B(w) = N(w): by the definition of m and Definition 5.8, B(m, w) = m(w) = N_w for any w. Since \( M = \{m\} \), by Definition 5.8 again B(w) = N(w). From this F and ≰ are easily shown to be pointwise equivalent over L_B.

\[ \square \]

Theorem C.6. Call a neighbourhood frame F = < W, N > truthful iff for each w, w ∈ \( \bigcap N(w) \).^5 For each methods frame ≰, there is a truthful neighbourhood frame F that is pointwise equivalent over L_K.

Proof. Let ≰ = < W, M^R, R > be any methods frame for K. Define the neighbourhood frame F = < W, N > such that for any w, N(w) = K(w) and for any model M in F, [Kφ]^M = {w : [φ]^M ∈ N(w)} . We prove as before that for any M in F and \( \forall \phi \) in ≰, \( \models^w_M Kφ \) iff \( \models^w_M Kφ \) and that the frames are pointwise equivalent. Moreover, we prove that for any w, w ∈ \( \bigcap N(w) = \bigcap K(w) \):

For any w, w ∈ \( \bigcap K(w) \) iff \( \forall m \forall p (p ∈ K(m, w) → w ∈ p) \) (Definition 5.8).

---

5. The class of truthful neighbourhood frames is the class of neighbourhood frames which validate the schema \( \forall \phi → \phi \). See Chellas (1980, 224).
For any \( w, m, p \), if \( p \in K(m, w) \) then \( m \in M^l \) and \( p \in m(w) \) (Definition 5.8).

if \( m \in M^l \) then \( p \in m(w) \rightarrow w \in p \) (Definition 5.7 and reflexivity of \( R \))

Thus for any \( w, m, p \), if \( p \in K(m, w) \) then \( w \in p \). So \( w \in \bigcap K(w) \) for any \( w \).

Conversely, let \( \mathcal{F} = < W, N > \) be any neighbourhood frame such that for any \( w, w \in \bigcap N(w) \). Define the methods frame \( \mathcal{G} = < W, M^B, R > \) such that \( M^B = \{ m \} \) where \( m \) is such that \( m(w, \pi) = N(w) \) for any \( w, \pi \) and \( R \) is identity. We prove that \( M = \{ m \} \) and \( B(m, w) = N(w) \) as before. Moreover, we prove that for any \( w, m \in M^l(w) \):

For any \( w : m \in M^l(w) \) iff for any \( w', p', wRw' \rightarrow (p' \in m(w') \rightarrow w' \in p') \) (Definition 5.7),

iff for any \( p', p' \in m(w) \rightarrow w \in p' \) (\( R \) is identity),

iff \( w \in \bigcap m(w) \),

iff \( w \in \bigcap N(w) \) (definition of \( m \)), which is true by assumption.

Thus \( m \in M^l(w) \) for any \( w \). Since \( M = \{ m \} \), it follows that \( K(w) = K(m, w) = N(w) \). From this we show as before that \( \mathcal{G} \) and \( \mathcal{F} \) are pointwise equivalent with respect to \( K \). \( \square \)

The results mean that the \( B \) and \( K \) schemas valid in the class of methods frames are just those valid in the class of neighbourhood frames and in the class of truthful neighbourhood frames, respectively (see section 5.5.1).

Given the equivalences of Theorems C.5 and C.6, why prefer methods models to simpler neighbourhood ones? Essentially, because methods models allows us to derive a set of facts that would be treated as primitive in a simple neighbourhood semantics models. A simple example: despite the equivalences C.5 and C.6, the class of methods frames is not the class of neighbourhood frames for two modalities \( < W, N^B, N^K > \) where the second neighbourhood function is truthful. For it is easy to see that in our models, \( p \in K(w) \rightarrow p \in B(w) \) for any \( w, p \) (Theorem 5.2), while we can construct neighbourhood models such that \( p \in N^K \wedge p \notin N^B \). Of course we could introduce the notion of belief-knowledge neighbourhood frames \( < W, N^B, N^K > \) such that at any \( w \), \( N^K(w) \subseteq N^B(w) \) (knowledge en-
tails belief) and $w \in \bigcap N^K(w)$ (knowledge entails truth). But that would amount to treating those facts as unexplained primitives. By contrast, those constraints on knowledge are derived in methods models from a definition of knowledge.

The same goes for other axioms. A much-discussed axiom for knowledge is 4, according to which knowing is knowing that one knows: $Kp \rightarrow KKp$. A neighbourhood frame validates 4 iff: $p \in N^K(w) \rightarrow \{w' : p \in N^K(w')\} \in N^K(w)$ for any $p, w$. The condition is a transparent restatement of axiom 4: if $p$ is among the propositions known at $w$, then so is the proposition that holds wherever $p$ is among the propositions known. The condition on the model does not shed any light on whether, when or why the axiom should hold. Thus we are left to decide directly on the basis of the axiom whether we think our agents would or should satisfy it. By contrast, in methods models, the axiom is derived from the psychological model of the agent and the transitivity of background alethic modality. If on the relevant sense of possibility, what is possibly possible is possible, we show that the agent satisfies axiom 4 for knowledge if it has a “confident introspection” method $m^{pc}$ such that, in non-formal terms: if she believes that $p$ out of $m$ then she believes that she knows that $p$ on the basis of a combination of $m$ and $m^{pc}$ (section 5.5.4). This gives a better grasp on how and when an agent is able to know that she knows. For a start, it shows that it is not a trivial affair: one can easily find counterexamples to the axiom for agents who do not introspect or for a non-transitive space of possibilities.

The explanatory advantages of methods models are due to the fact that they contain more structure than neighbourhood ones. The neighbourhood functions $B$ and $K$ are not given as primitives, but constructed out of a set of methods. The construction gives us an insight into a structure of $B$ and $K$ that is not simply reducible to the structure of the set of propositions they map to (the structure of $\{p : p \in B(w)\}$ and $\{p : p \in K(w)\}$ at each $w$). The additional structure is reflected in the methods operators $B\mu : \phi$ and $K\mu : \phi$. For instance, we get validities such as:
\[ \models^{\mathfrak{M}} K\mu : (\phi \rightarrow \psi) \rightarrow (K\nu : \phi \rightarrow Km^D \circ (\mu + \nu) : \psi) \]

\[ \models^{\mathfrak{M}} K\mu : \phi \rightarrow Km^D \circ \mu : B\phi \]

which cannot be stated with the unary operators.

### C.3 Counterexamples to M, N, K and 4

**Example C.1.** Counterexample to $M_B$ and $M_K$. We construct a model where the agent believes and knows that $p \land q$, but does not believe or know that $q$. Consider a frame $\mathfrak{F} = \langle W, M^B, R \rangle$ where $W = \{w_1, w_2\}$ with $p = \{w_1\}$, $q = W$ and $M^B = \{m\}$ where $m$ is such that $m(w_1, \pi) = \{p\}$ and $m(w_2, \pi) = \emptyset$ for any $\pi$, and $R$ any reflexive accessibility relation. Consider $\mathcal{M}$ in $\mathfrak{F}$ such that $V(p) = p$ and $V(q) = q$. It is easy to check that $M = \{m\}$ and that $\models^{\mathfrak{F}} M_B(p \land q)$ but $\not\models^{\mathfrak{F}} M_Bq$, and similarly for $K$ since however $R$ is defined, the method $m$ is infallible at $w_1$ ($m \in M^I(w_1)$).

\[ p = p \land q \quad q \]

![Figure C.1: Counterexample to $M_B$ and $M_K$.](image)

(The illustrations are explained p. 239 above.)

The methods frames $\mathfrak{F}$ constructed from a neighbourhood frame $\mathcal{F} = \langle W, N \rangle$ following the procedure in Theorem C.6 are such that the agent’s unique method is infallible and so they validate $K\phi \leftrightarrow B\phi$. (They are “excellent agent frames”, in our terminology: see Definition 5.18). Consequently the counterexamples to the K schemas built this way, like Example C.1, all involve a failure of belief, and a failure of the corresponding B schema. However there are two ways for knowledge to fail
in methods models: failure of belief, but also fallibly-based belief. It will be instructive to look at two examples of the latter.

Example C.2. Counterexample to $N_K$. The agent has a method that leads her to believe both the tautology and a false proposition. Though the agent believes the tautology ($W$), she fails to know it, because her belief is fallibly based.

Consider $\mathfrak{f} = \langle W, M^B, R \rangle$ where $W = \{w_1, w_2\}$, with $p = \{w_2\}$, and $M^B = \{m\}$ where $m$ is s.th. $m(w_1, \pi) = \{W, p\}$ and $m(w_2, \pi) = \{W\}$ for any $\pi$; $R$ is any reflexive accessibility relation. Consider $M$ in $\mathfrak{f}$ such that $V(p) = p$. Since $p \in m(w_1)$ but $w_1 \not\in p$, $m \not\in M(I(w_1))$. From this it follows that $\models_{w_1} B\top$ but $\not\models_{w_1} K\top$ (Definitions 5.8 and 5.10).

\begin{figure}[h]
\centering
\begin{tikzpicture}
  \node (w1) at (0,0) {$w_1$};
  \node (w2) at (1.5,0) {$w_2$};
  \node (w3) at (3,0) {$w_3$};
  \node (w0) at (4.5,0) {$(w_0)$};
  \node (w4) at (0,1.5) {$w_1$};
  \node (w5) at (1.5,1.5) {$w_2$};
  \node (w6) at (3,1.5) {$w_3$};

  \draw [->] (w1) to node [midway, right] {$m$} (w2);
  \draw [->] (w2) to node [midway, right] {$m$} (w1);
  \draw [->] (w1) to node [midway, above] {$m$} (w4);
  \draw [->] (w2) to node [midway, above] {$m$} (w5);
  \draw [->] (w3) to node [midway, above] {$m$} (w6);
  \draw [->] (w4) to node [midway, above] {$m$} (w0);
  \draw [->] (w5) to node [midway, above] {$m$} (w0);
  \draw [->] (w6) to node [midway, above] {$m$} (w0);

\end{tikzpicture}
\caption{Counterexample to $N_K$.}
\end{figure}

Example C.3. Counterexample to $K_K$. We consider an agent that believes $p \rightarrow q$ and $p$ by infallible methods. The agent also believes $q$, but on the basis of a method that would lead her to believe $q$ even if it was false, so the agent fails to know that $q$.

Consider $\mathfrak{f} = \langle W, M_B, R \rangle$ where $R = W \times W$ and $W = \{w_1, w_2, w_3\}$ with $p = \{w_1\}$ and $q = \{w_1, w_2\}$. $M^B = \{m, n, r\}$ such that $m(w, \pi) = \{p\}$ and $n(w_2, \pi) = n(w_3, \pi) = \emptyset$ for any $\pi$, and $r(w, \pi) = \{q\}$ for any $w, \pi$.

Consider $\mathfrak{m}$ in $\mathfrak{f}$ such that $V(p) = p$ and $V(q) = q$. We have $\models_{w_1} p \rightarrow q = W$. Since $q \in r(w_3)$ but $w_3 \not\in q$, and since $w_1 R w_3, r \not\in M^I(w_1)$ (by Definition 5.7). It follows that at $w_1$ we have: $\models_{w_1} K(p \rightarrow q), \models_{w_1} Kp, \models_{w_1} Bq$ and yet $\not\models_{w_1} Kq$. 
Example C.3 models our Watson case. Though the agent knows two propositions that together entail \( q \) (namely \( p \to q \) and \( p \)), and though she believes that \( q \), her belief that \( q \) is not based on deduction from the two others; rather, it is has an independent and unreliable basis that would lead her to still believe that \( q \) without knowing that \( p \) and \( p \to q \), and even if \( q \) was false.

Example C.4. Counterexample to \( 4_K \). The agent has a method that leads her to believe that she knows that \( p \), even at a world where she does not.

Consider \( \mathcal{F} = \langle W, M^B, R \rangle \) where \( W = \{w_1, w_2\} \), with \( p = \{w_1, w_2\} = W \), \( q = \{w_1\} \) and \( M^B = \{m, n\} \) where \( m \) is s.th. \( m(w_1, \pi) = \{p\} \) for any \( \pi \), and \( n \) is s.th. \( n(w, \pi) = \{q\} \) for any \( w, \pi \), and \( R = W \times W \).

Consider \( \mathcal{M} \) in \( \mathcal{F} \) such that \( V(p) = p \). Since \( p \in m(w_1) \) and \( m \in M^I \), \( \Vdash_{w_1}^{\mathcal{M}} KP \). Since \( \models_{w_1}^{\mathcal{M}} q \) and \( q \in n_{w_1}, \models_{w_1}^{\mathcal{M}} BKp \). But since \( w_1Rw_2, q \in n(w_2) \)
and \( q \not\in w_2, n \not\in M^l(w_1) \), and since there is no other method \( r \) such that \( q \in r(w_1) \), we have \( \not\models_{M_{w_1}}^R \text{KKp} \). (Definitions 5.8 and 5.10).

C.4 Further exploration of Reasoning methods

Axioms \( M \) and \( C \) for belief and knowledge follow from axiom \( K \). But it is also possible to get them separately, by defining the two following methods:

Definition C.3. Single-Premise Deduction is the method \( m^{SD} \) s.th \( m^{SD}(w, \pi) = \{ p : \exists q \in \pi(q \subseteq p) \} \) for any \( w, \pi \).

Conjunctive Deduction is the method \( m^{CD} \) such that \( m^{CD}(w, \pi) = \{ p : \exists q, r \in \pi(p = q \cap r) \} \) for any \( w, \pi \).

Theorem C.7. For any methods model \( M \), if \( m^{SD} \in M \), then \( (M_B) \models \text{B}(\phi \land \psi) \rightarrow (\text{B}\phi \land \text{B}\psi) \) and \( (M_K) \models \text{K}(\phi \land \psi) \rightarrow \text{K}(\phi \land \psi) \) for any \( \phi, \psi \).

For any methods model \( M \), if \( m^{CD} \in M \), then \( (C_B) \models \text{B}(\phi \land \psi) \rightarrow \text{B}(\phi \land \psi) \) and \( (C_K) \models \text{K}(\phi \land \psi) \rightarrow \text{K}(\phi \land \psi) \).

Proof. The proofs are analogous the proof of \( K_B \) and \( K_K \) (Theorem 5.7).

For \( (M_B) \) and \( (M_K) \), assume that \( \mathcal{M}, w \) are s.th. \( \models_{M_{w}}^R \text{B}(\phi \land \psi) \) and that \( m^{SD} \in M \). Then there are \( p, q \) such that \( [\phi]_{M_{w}}^R = p, [\psi]_{M_{w}}^R = q \), and \( p \cap q \in m(\omega) \) for some \( m \in M \). We show that \( p, q \in (m^{SD} \circ m)(w) \), that \( m^{SD} \circ m \) is infallible if \( m \) is, and that \( m^{SD} \circ m \in M \). From this \( (M_B) \) and \( (M_K) \) follow.

For \( (C_B) \) and \( (C_K) \), assume that \( \mathcal{M}, w \) are s.th. \( \models_{M_{w}}^R \text{B}\phi \land \text{B}\psi \) and that \( m^{CD} \in M \). Then there are \( p, q \) such that \( [\phi]_{M_{w}}^R = p, [\phi]_{M_{w}}^R = q \), and \( p \cap q \in m(\omega) \) and \( q \in m^n(\omega) \) for some \( m, n \). We show that \( p \cap q \in (m^{CD} \circ (m + n))(w) \), that \( m^{CD} \circ (m + n) \) is infallible if \( m \) and \( n \) are, and that \( m^{CD} \circ m \in M \). From this \( (C_B) \) and \( (C_K) \) follow. \( \Box \)

The relation between Multi-Premise Deduction, Single-Premise Deduction and Conjunctive Deduction is straightforward:

Corollary C.1. \( m^D = m^{SD} \circ m^{CD} \).

Proof. Evident from Definitions 5.12 and C.3. \( \Box \)
In neighbourhood models, axioms $M$, $C$ and $K$ have been correlated to corresponding properties of the topology of sets of sets (see Chellas, 1980, 215–216). A set of sets $S \subseteq \mathcal{P}(W)$ is supplemented or closed under supersets iff $\forall X, Y \subseteq W((X \in S \land X \subseteq Y) \rightarrow Y \in S)$, is closed under finite intersections iff $\forall X, Y \in S(X \cap Y \in S)$, contains its core iff $\bigcap S \in S$, and is augmented iff it is supplemented and contains its core. We say that a neighbourhood function $W \rightarrow \mathcal{P}(W)$ is supplemented, closed under finite intersections, and augmented iff it maps to supplemented, closed under finite intersections, and augmented sets, respectively. Supplemented neighbourhood functions satisfy $M$, neighbourhood functions that are closed under finite intersections satisfy $C$, and augmented ones satisfy $K$. It can be shown that the methods $m^{SD}$, $m^{CD}$ and $m^D$ ensure that the $B$ and $K$ neighbourhood functions are respectively supplemented, closed under finite intersections, and augmented. We do not present the proofs here. The reader will easily construe them by considering, for a given method $m$, the series of combined methods $m^{Dm_k}$, $k \in \mathbb{N}$ such that $m^{Dm_0} = m$, and $m^{Dm_k} = m^D \circ m^{Dm_{k-1}}$ for any $k \geq 1$, and analogous series for $m^{SD}$ and $m^{CD}$.

Satisfying the knowledge axioms $M_K$, $C_K$, $K_K$ does not guarantee satisfaction of the corresponding belief axioms, and conversely. It is in principle possible that an agent believes the logical consequences of what she believes on the basis of infallible methods but does not believe all the logical consequences of what she believes on the basis of fallible methods.

Finally, note that we have only stated sufficient conditions for a methods frame to satisfy $M$, $C$, $K$ and $N$, not necessary ones. There are frames that validate those schemas for belief and/or knowledge without Deduction and Pure Reason. Consider two examples:

- $M_B = \{m^K\}$. The agent believes and knows the tautology, and only the tautology, at any world. Trivially, the agent validates $K_B$ and $K_K$, yet she does not have $m^D$.

- Suppose that an agent is such that whenever, for some $w$, $m$, $n$, $p \in B(m, w)$ and $(W \setminus p) \cup q \in B(n, w)$, there is some third method $r$ such that $q \in B(r, w)$, yet the third method is not the result of
C.4. Further exploration of Reasoning methods

combining \( m \) and \( n \) with \( m^D \). For instance, one may assume that \( r \) also outputs some true propositions (say, \( \{w\} \) at any \( w \)) that do not follow from the outputs of \( m \) and \( n \). Such an agent can satisfy \( K_B \) and/or \( K_K \) without having \( m^D \).

Therefore the methods \( m^R \) and \( m^D \) fail to identify the class of methods frames that validate the schema of normal modal logics (\( KN \)). By contrast, in neighbourhood semantics, these class of frames can be identified as the ones in which the neighbourhood function is augmented. Is that a defect of our models? Quite the contrary. An agent may satisfy the \( KN \) schemas “accidentally”, so to speak. Imagine an agent that forms beliefs by listening to various people’s testimonies. Suppose that whenever the agent has heard \( p \) and \( p \rightarrow q \) from some persons, there happens to be, by sheer coincidence, a person that tells her that \( q \). The agent thereby satisfies \( K_B \), \( B(p \rightarrow q) \rightarrow (Bp \rightarrow Bq) \). Yet that is intuitively accidental, because her belief that \( q \) is unconnected to her believing \( p \) and \( p \rightarrow q \). By contrast, it is not all accidental that an agent satisfies \( K_B \) if the agent has the Deduction method, because the method ensures that she has a belief that \( q \) based on her having beliefs that \( p \) and that \( p \rightarrow q \). The upshot is that, far from being a deficiency of methods frames, the fact that there is no natural class of methods frames that validates \( KN \) rather shows that the epistemic and doxastic \( KN \) are superficial rather than deep generalities about knowledge and belief. For instance, the deep generality behind \( K \) is that a (certain idealised type of) agent deduces all the logical consequences of what she knows, and that generality can only be stated in the more complex language that allows reference to methods (Definition 5.10), as we pointed out (see Theorem 5.8, p. 260).
C.5 Belief, knowledge and information

As understood here, information is an objective notion of content. The information contained in a proposition is just the set of possibilities compatible with it: namely, the set of worlds in which it is true. We define information for a set of propositions as the set of worlds that are compatible with each proposition in the set. As a special case, the empty set of propositions is compatible with any world: its information is the same as that of a tautology. A set of propositions containing a contradiction or contradictory propositions is not compatible with any world, so its information is the empty set. With this notion we can define the (conditional or unconditional) information provided by a method at a world as the set of possibilities compatibles with the (conditional or unconditional) outputs of that method at that world, and the agent’s doxastic and epistemic information as the set of possibilities compatible with what an agent believes and with what she knows, respectively.

**Definition C.4.** For any set of propositions \( \pi \), we define:

\[
I(\pi) = \{ w \in W : \forall p (p \in \pi \rightarrow w \in p) \}.
\]

For any method \( m \), world \( w \) and set of premises \( \pi \):

\( I(m(w, \pi)) \) is the information provided by \( m \) at \( w \) on the basis of \( \pi \).

For any methods model \( \mathcal{M} \) and world \( w \):

\( I(B(w)) \) is the agent’s doxastic information at \( w \), and

\( I(K(w)) \) is the agent’s epistemic information at \( w \).

Note that the information of a method is defined independently of whether the agent has it; by contrast, doxastic and epistemic information are agent- and model-dependent.

The relations between \( I(m(w, \pi)) \) and \( I(B(w)) \) and \( I(K(w)) \) are straightforward. The agent’s doxastic information is the unconditional information given by all the methods she has: \( I(B(w)) = \bigcap_{m \in \mathcal{M}} I(m(w)) \), provided

6. Revision note. The definitions in this appendix have been amended to take into account methods with no outputs. We have also generalised the notion of information to methods the agents do not have and simplified some results.
that $M$ is not empty. (If $M$ is empty, $I(B(w)) = W$.) The agent’s epistemic information is the unconditional information given by the infallible methods she has: $I(K(w)) = \bigcap_{m \in M \cap M'(w)} I(m(w))$, provided $M \cap M'(w)$ is not empty, otherwise $I(K(w)) = W$.

The formal representation of such notions is familiar from Hintikka (1962) models. The worlds compatible with what I believe are just those worlds where every proposition I believe holds, and similarly for knowledge. Our notions of information thus correspond to standard Kripke models:

**Definition C.5.** For any frame $\mathfrak{F}$, let $R^B, R^K \subseteq W \times W$ be such that, for any $w, w'$:

- $wR^B w'$ iff $w' \in I(B(w))$,
- $wR^K w'$ iff $w' \in I(K(w))$.

Each of $\langle W, R^B \rangle$, $\langle W, R^K \rangle$ is a Kripke frame. We can introduce corresponding modal operators $D$ and $E$ in the language. For any methods model $\mathfrak{M}$,

$$\llbracket D \phi \rrbracket^{\mathfrak{M}} = \{ w : \forall w' (wR^B w' \rightarrow w \in I(w)) \},$$

$$\llbracket E \phi \rrbracket^{\mathfrak{M}} = \{ w : \forall w' (wR^K w' \rightarrow w \in I(w)) \}.$$ 

**Corollary C.2.** For any methods models, $B \phi \rightarrow D \phi$ and $K \phi \rightarrow E \phi$.

**Proof.** Let $\mathfrak{M}, w$ be such that $\mathfrak{M}, w \models B \phi$. By Definitions 5.10 and 5.8, $[\phi]^{\mathfrak{M}} \in B(w)$. By Definition C.4, $I(B(w)) \subseteq [\phi]^{\mathfrak{M}}$, and by Definition C.5, $\forall w' (wR^B w' \rightarrow w \in [\phi]^{\mathfrak{M}})$, so $\mathfrak{M}, w \models D \phi$. And similarly for $K \phi \rightarrow E \phi$. □

By contrast, it easy to find models in which $D \phi \rightarrow B \phi$ fails: for instance, $D \top$ holds at any model, but $B \top$ fails at some. This gives a sense in which $D$ and $E$ represent the information contained in one’s belief and knowledge: explicitly, when $D \phi \land B \phi$, and implicitly, when $D \phi \land \neg B \phi$, and similarly for $E$ and $K$.

Intuitively, a method is informative iff it can reduce the set of possibilities the agent considers or should consider. Formally, we define:
**Definition C.6.** A method \( m \) is **uninformative** iff \( I(m(w, \pi)) \supseteq I(\pi) \) for all \( w, \pi \). A method is informative iff it is not uninformative.  

In particular, \( m \) is uninformative only if its unconditional outputs are not more informative than the tautology: \( I(m(w)) \supseteq I(\emptyset) = W \) for any \( w \). That is, at any world, either \( m \) unconditionally outputs nothing or the tautology.

**Corollary C.3.** If \( m \) is an uninformative method, then for any \( n, w, \pi \):

\[
I(m + n(w, \pi)) \supseteq I(\pi \cup n(w, \pi)),
\]

\[
I(m \circ n(w, \pi)) \supseteq I(n(w, \pi)).
\]

Proof. Note first that \( I(\pi \cup \pi') = I(\pi) \cap I(\pi') \) for any \( \pi, \pi' \). By Definition C.4, \( w' \in I(\pi \cup \pi') \) iff \( \forall p ((p \in \pi \lor p \in \pi') \rightarrow w' \in p) \), or equivalently, \( \forall p (p \in \pi \rightarrow w' \in p) \land \forall p (p \in \pi' \rightarrow w' \in p) \), that is \( w' \in I(\pi) \cap I(\pi') \).

From this and Definition 5.2, \( I(m + n(w, \pi)) = I(m(w, \pi)) \cap I(n(w, \pi)) \) for any \( m, n, w, p \). Now if \( m \) is uninformative, \( I(m(w, \pi)) \supseteq I(\pi) \) (Definition C.6), so \( I(m(w, \pi)) \cap I(n(w, \pi)) \supseteq I(\pi) \cap I(n(w, \pi)) \).

By Definition 5.2, \( m \circ n(w, \pi) = m(w, n(w, \pi)) \) for any \( m, n, w, \pi \). And if \( m \) is uninformative, \( I(m(w, n(w, \pi))) \supseteq I(n(w, \pi)) \) (Definition C.6).  

The union of a uninformative method \( m \) with a method \( n \) does not give more information than is contained in premises or given by \( n \). The application of an uninformative method \( m \) to \( n \) does not give more information than \( n \) did not already give. Union and combinations of uninformative methods are uninformative:

**Corollary C.4.** If \( m \) and \( n \) are uninformative methods, \( m + n \) and \( m \circ n \) are uninformative.

Proof. Suppose \( m \) and \( n \) are uninformative. By Corollary C.3, \( I(m + n(w, \pi)) \supseteq I(\pi \cup \pi) = I(\pi) \) for any \( w, \pi \). By Corollary C.3 again, \( I(m \circ n(w, \pi)) \supseteq I(n(w, \pi)) \) for any \( w, \pi \), and since \( n \) is uninformative, \( I(n(w, \pi)) \supseteq I(\pi) \).  

---

7. Thanks to Timothy Williamson here.
Uninformative methods are guaranteed to preserve truth, in the following sense:

**Corollary C.5.** Say that a method \( m \) is truthful at \( w \) iff \( w \in I(m(w)) \). If a method \( m \) is uninformative, then for any \( n \), \( m + n \) and \( m \circ n \) are truthful at \( w \) if \( n \) is.

**Proof.** Suppose \( n \) is truthful at \( w \): \( w \in I(n(w)) \). If \( m \) is uninformative, \( I(m + n(w)) \supseteq I(n(w)) \) (Corollary C.3), so \( w \in I(m + n(w)) \). Again, if \( m \) is uninformative, \( I(m \circ n(w)) \supseteq I(n(w)) \) (Corollary C.3), so \( w \in I(m \circ n(w)) \). □

This characterises the sense in which uninformative methods are risk-free. Uniting an uninformative method with another or applying it to another cannot lead to false beliefs unless the original method did.

**Theorem C.8.** Deduction and Pure-Reason are uninformative.

**Proof.** By Definition 5.11, \( m^R(w, \pi) = \{W\} \) for every \( w, \pi \). So \( I(m^R(w, \pi)) = W \supseteq I(\pi) \) for any \( w, \pi \).

By Definition 5.12, \( m^D(w, \pi) = \{p : \exists q, r \in \pi(p \supseteq q \cap r)\} \). Suppose \( w' \in I(\pi) \) for some \( \pi \); then \( w' \in q \cap r \) for any \( q, r \in \pi \) (from Definition C.4), so \( w' \in I(m^D(w, \pi)) \) at any \( w \). □

**Theorem C.9.** Say that a method \( m \) is included in a method \( n \) iff for all \( w, \pi \), \( m(w, \pi) \subseteq n(w, \pi) \). The most inclusive uninformative method is the Total Reasoning method \( m^{TR} \) such that \( m^{TR}(w, \pi) = \{p : p \supseteq I(\pi)\} \) for any \( w, \pi \).

**Proof.** Let \( m \) be a method not included in \( m^{TR} \). Then there is some \( p, w, \pi \) such that \( p \in m(w, \pi) \) and \( p \not\supseteq I(\pi) \). So for some \( w', w' \in I(\pi) \) but \( w' \notin p \). Hence \( w \notin I(m(w, \pi)) \) and \( I(m(w, \pi)) \not\supseteq I(\pi) \), so \( m \) is informative. □

By contrast, Introspection methods are typically informative.

**Theorem C.10.** Introspection and Confident Introspection are informative. For some methods models, there are \( w, \pi \) such that \( I(m^{\pi i}(w, \pi)) \not\supseteq I(\pi) \), and similarly for \( m^{\pi c} \) and \( m^{\pi c} \).

**Proof.** Model for \( I(m^{\pi i}(w, \pi)) \not\supseteq I(\pi) \). Let \( W = \{w_1, w_2\} \), \( p = \{w_1\} \), \( q = \{w_2\} \) and \( M = \{m, m^{\pi i}\} \) where \( m \) is such that for any \( \pi \), \( m(w, \pi) = \{W\} \) if \( w = w_1 \)
and $\emptyset$ otherwise, and $m^{pi}$ is such that for any $w$: $m^{pi}(w, \{W\}) = \{w_1\} = \{p\}$, $m^{pi}(w, \{p\}) = \{w_1\} = \{p\}$, $m^{pi}(w) = m^{pi}(w, \{q\}) = \emptyset$.

It is easy to check that $m^{pi}$ is a positive Introspection method: for any $\pi$, $m^{pi}(w, \pi)$ is the set of propositions $b_r$ such that $r \in \pi$ and $b_r = \{w' : \exists m (m \in M \land r \in m(w'))\}$. (In particular, $\{w_1\}$ is the set of all worlds where some method (e.g. $m^{pi} \circ m$, or $m^{pi} \circ m^{pi} \circ m$, ...) maps to $p$, and since $\{w_1\} = p$, it is also the set of all worlds where some method maps to the set of all worlds where some method maps to $p$, and so on.) At $w_1$ we have: $I(\{W\}) = W$ but $I(m(w_1, \{W\})) = I(\{p\}) = p$. So $m^{pi}$ is informative. Similar models can be built for $m^{ni}$, $m^{pc}$, $m^{nc}$.

The results are fairly intuitive. Typically, $p$ and $Bp$ do not hold at the same worlds. For that reason, an Introspection method that “adds” a belief that $Bp$ wherever the agent believes that $p$ typically narrows down the sets of worlds compatible with the agent’s belief. Similarly, $p$ and $Kp$ do not typically hold at the same worlds. That is why Introspection and Confident Introspection are informative methods. Because they narrow down the set of worlds compatible with the agent’s beliefs, they are not risk-free methods: it may be true that $p$ and false that $Bp$, or true that $Bp$ and false that $Kp$. They are inductive methods, as we defined them section 5.3.4. Correlatively, the axioms of epistemic logic that rely on them (4 and 5) are not a matter of pure rationality or inner coherence; they require reliable information-gathering methods.
Appendix D

Safety, Closure and Knowledge of the Past

One advertised advantage of a well-formulated safety condition on knowledge is its compatibility with the intuitive idea that knowledge is closed under competent deduction (Williamson, 2000, 117, 2009a, 9–14; Sosa 1999b, 149). However, safety is also commonly taken to be time-relative (Sainsbury, 1997, 913; Peacocke 1999, 319; Williamson, 2000, 124, 2009a, 21). In this section I argue that under natural assumptions, and unless the safety condition is made implausibly weak, the two claims are inconsistent. Safety conditions face structurally similar difficulties with testimony and memory, but I will only point them out without exploring them further.

D.1 Time-relative safety

The ordinary notions of safety and danger are time-relative. Having caught the last flight out of a besieged city, I am now safe from being shot. Yet if I nearly missed it, it was not the case earlier on that I was safe from being shot (Williamson, 2000, 124).

Time-relative safety is naturally modelled by a branching structure. Let \( H_{t,w} \) state the history of world \( w \) up to and including time \( t \), and let \( C_w \) be a closeness constraint (of which more shortly):

**Closeness (C)** A world \( w' \) is close to \( w \) at \( t \) iff \( H_{w,t} \& C_w \) holds in \( w' \).
Let $F(w, t)$, the fan of $w$ at $t$, be the set of worlds close to $w$ at $t$. We say:

**Safety (S)** At $w$, $p$ is safe from occurring at $t$ iff $p$ fails in all worlds in $F(w, t)$.

Almost nothing would be safe from occurring if any world that merely had the same past as actuality was counted as close. To avoid the trivialisation, some constraint $C_w$ is needed to exclude the irrelevant worlds among those who share their initial history with $w$. The laws of $w$ are an obvious candidate. However, $C_w$ may be stronger than the laws in some respects and weaker in others. Stronger: if chanciness is widespread, some wild low-chance outcomes of chancy processes may have to be excluded from the close worlds, even if they are compatible with the laws and the past (Hawthorne and Lasonen-Aarnio, 2009, 94). Weaker: if the laws are deterministic, some worlds with the same past but a different future may be included, even though they require some small miracles by the lights of the actual laws. Otherwise anything that does not occur would be safe from occurring, another unpalatable result. (Though one has to grant some pull for saying that in deterministic worlds, anything that did not happen was in fact safe from occurring.) Finally, however $C_w$ is defined, it must at least hold in $w$, so that each world is close to itself at all times.

Peacocke (1999, 315) holds that worlds close to $w$ must satisfy the laws of $w$. To avoid trivialising safety in deterministic settings, he rejects the idea that all worlds close to $w$ at $t$ share their history with $w$ up to $t$. The resulting view is counter-intuitive. Surely, if I am in the plane, I am now safe from having missed the flight, though I may have been in danger of missing it at some time earlier. And surely, if I am not shot now, I am now safe from being shot now, though I may not now be safe from being shot later. These intuitive inferences are not valid, however, if there are worlds that are close now in which I have not taken the plane or in which I am now shot.

I will say that an expansion occurs if some worlds that were not close at a time become close at a later time, and that contraction occurs if some worlds that were close at a time are not close at a later time:
Expansion and contraction. If $t_2$ is later than $t_1$, an expansion occurs at $w$ between $t_1$ and $t_2$ iff some world in $F(w, t_2)$ was not in $F(w, t_1)$, and a contraction occurs at $w$ between $t_1$ and $t_2$ iff some world in $F(w, t_1)$ is not in $F(w, t_2)$.

(As defined, expansion and contraction are not exclusive of each other.)

Say that a world is bergsonian iff some expansion occurs in it. Bergsonian worlds are excluded by (C) and the uncontroversial assumption that:

**History is cumulative (HC).** For any $w$, if $t_2$ is later than $t_1$, $H_{w, t_2}$ entails $H_{w, t_1}$.

For suppose $w^* \in F(w, t_2)$ and $t_2$ is later than $t_1$. By (C), $H_{w, t_2} \& C_w$ holds in $w^*$. By (HC), $H_{w, t_1}$ holds in $w^*$. By (C) again, $w^* \in F(w, t_1)$. So $F(w, t_2) \subseteq F(w, t_1)$: whatever is not safe from occurring at a time was not safe from occurring at any earlier time.

There are considerations in favour of bergsonian worlds, however. In deterministic settings, we may want to consider as close at $t$ worlds in which small miracles occur but only slightly after $t$. At some time $t_2$, for instance, we may have a close world in which a small miracle occurs after $t_2$; but that world cannot have been close at an earlier time $t_1$, for it has a miracle elsewhere than right after $t_1$. So the actual world is bergsonian.

On such views, a time-relative constraint $C_{w,t}$ should be substituted for $C_w$ in (C). Then the fact that $H_{w, t_1} \& C_{w, t_2}$ holds at $w^*$ does not entail that $H_{w, t_1} \& C_{w, t_1}$ holds at $w^*$, which blocks the proof above.

### D.2 Safety and knowledge of the past

On safety accounts of knowledge, one knows only if one is relevantly safe from error. The epistemically relevant notion of safety need not exactly match the general-purpose notions of safety and danger, but it should at least have analogous structural features (Williamson, 2009a, 9).

---

1. Bergson (1934) thought that the realm of the possible increased with time.
Here we assume that the relevant features are the ones made explicit by (C), (S) and (HC): closeness and a branching time structure.

One’s belief that \( p \) is safely true iff “one could not easily have been wrong in a similar case” (Williamson, 2000, 147). At first sight, we may gloss the idea as follows:

**Present-indexed belief safety (BS1).** One’s belief that \( p \) is safely true at \( w, t \) iff at \( w, t \) my having a relevantly similar but mistaken belief at \( t \) was safe from occurring.\(^2\)

(BS1) gives intuitive results for knowledge of the future. Let \( w \) be the actual world. At \( t \), Alice holds a ticket in a fair lottery which will be drawn the next day. I form the true belief that her ticket is a loser. Under natural assumptions there is a world \( w' \) in \( F(w, t) \) in which she wins and I have formed the same belief, which is relevantly similar. Thus I am not safe from forming a relevantly similar false belief. A safety condition thus delivers the intuitive verdict that I do not know.

However, (BS1) is too weak concerning knowledge of the past and the present. Suppose the lottery has been drawn and Alice lost, though the results have not been announced yet. Then \( H_{w,t} \) implies that the lottery has been drawn and that she lost. By (C), there are no worlds in \( F(w, t) \) in which I form the false belief that her ticket is a loser. By (S), I am now safe from mistakenly believing that it is a loser. Furthermore, nothing ensures that there is some other lottery ticket about which I am now in danger of forming a relevantly similar belief that it is a loser. Suppose that the ground for my belief is that Alice is exceptionally unlucky: then I am not in danger of forming the belief that anyone else’s ticket is a loser. Of course, if Alice could have bought another ticket, I was earlier on in danger of forming a similar belief about that other ticket. But since she has not bought it, I am not now in danger of forming that belief either.

---

\(^2\) A belief is mistaken iff it is false or truth-valueless (Hawthorne, 2004, 56n). The precision that the mistaken belief must be had “at \( t' \)” should in principle be dropped and understood (in a relaxed form that allows for slightly different times) as implicit in the requirement that the belief be relevantly similar. However, it is useful to have it there to avoid any confusion between the time of safety from error and the time of belief possession, which come apart in (BS2) below.
Finally, there is no reason to suppose that some other false belief I am now in danger of forming will be relevantly similar to my lottery belief. In fact, the case may be stipulated so that I am not now in danger of forming any false belief. My past lottery belief will turn out safe by the standards of (BS1). Yet it is intuitively just as unsafe from error as my lottery belief in the future case.

A natural diagnosis is the following. Even though Alice could not win the lottery now, she could have won it. And similarly, even though I could not now wrongly believe she lost, I could have. That is, even though there is no world close now where I do, there was one earlier on. And that makes my belief unsafe in the relevant sense. Let \( p \) be the proposition that Alice’s ticket is a loser. Let \( t_p \) be the relevant time for \( p \); in this particular case, the time at which the drawing process begins. (We return to the notion shortly.) We say:

**Past-indexed belief safety (BS2).** One’s belief that \( p \) is safely true at \( w, t \)

iff at \( w, t_p \) my having a relevantly similar but mistaken belief at \( t \)

was safe from occurring.

On this proposal, the set of close worlds relevant to whether one knows some proposition \( p \) at \( t \) **varies with the target proposition \( p \) itself.** That ultimately explains why closure fails, as will become clear.

What is the time relevant for a given proposition? To simplify matters, assume that time is not dense. \(^3\) Consider now the **last unsettled time** (if any) of a proposition, that is the last time (if any) such that forever after the proposition is safe from having another truth-value from the one it actually has:

**Settling.** \( p \) settles in \( w \) iff there is a time \( t \) such that: there are both \( p \)-worlds

\(^3\) If time is dense, we cannot assume that a settling proposition \( p \) has a last unsettled time, just as there is no last number smaller than 2. We can only be sure that it either has a last unsettled time \( (u) \) or a first settled time \( (s) \). In the first case, we apply (BS2) with \( t_p = u \), as in the main text. In the second, we say that the corresponding belief is safely true iff there is a prior time \( s − \delta \) such that (BS2) holds for any \( t_p \) such that \( s − \delta ≤ t_p ≤ s \). In intuitive terms, the period \((s − \delta, s)\) is an ultimate period of unsettledness; instead of requiring that a mistake was safe from occurring at the last unsettled time, we require that a mistake was safe from occurring at some ultimate period of unsettledness. But such refinements are not relevant to our present concerns.
and \( \neg p \)-worlds in \( F(w, t) \) and for all \( t' \) later than \( t \), there are either only \( p \)-worlds in \( F(w, t') \), or only \( \neg p \)-worlds. In the former case we say that \( p \) **settles true after** \( t \) at \( w \); in the latter we say that \( p \) **settles false after** \( t \) at \( w \). We call \( t \) the **last unsettled time** for \( p \) at \( w \).  

Call **first truths** propositions that are never unsettled: for any \( t \), there are only \( p \)-worlds in \( F(w, t) \). Call **last truths** truths that are never settled: for any \( t \), there is some later \( t' \) with both \( p \)-worlds and \( \neg p \)-worlds in \( F(w, t') \). A proposition can be true at a world without ever settling: imagine an infinite sequence of heads with a close tails-world at each throw. Whether the laws are first or last truths depends on whether some close worlds have small miracles.

In the past lottery case, the last unsettled time is the relevant one. But one may worry that that time is too late when knowledge of the future is concerned. For if \( p \) settles in a quite distant future, safety will be evaluated only on the basis of worlds that share their history with actuality up to then: but what if the relevant error possibilities are much closer to the present time? The worry is misguided, however. Suppose that at \( w, t \) I believe that \( p \) and \( p \) is not settled. Let the relevant time for safety \( t_p \) be the last unsettled time for \( p \). Since \( p \) is unsettled at \( t_p \), we have a world \( w^* \) in \( F(w, t_p) \) in which \( p \) is not true. Since \( t_p \) is later than \( t \), \( H_{w,t} \) holds at \( w^* \). Since \( H_{w,t} \) holds at \( w^* \), \( w^* \) is a world in which I believe \( p \) at \( t \) and I have acquired the belief just as I actually have. (I assume here that the content of the belief and the way in which it is acquired do not depend on its future; more on this below.) That suffices to make it relevantly similar to my actual belief. So \( w^* \) is a world in which I have a relevantly similar but mistaken belief. My belief is not safely true by (BS2), and thus I fail to know. So however distant the last unsettled time is, it will not result in a weak safety condition: to the contrary, any settling time later than the time of belief formation ensures that the belief is unsafe. Conversely (and again provided that my belief is not future-dependent), I know that

---

4. If the world is bergsonian, some propositions are safe from being false at some time without being safe from being false at some later time. Hence the need to ensure that the proposition is **forever after** safe from being false.
There are exceptions to the conclusion of the last paragraph. If there was backward causation, the way a belief is acquired could depend on the future: I could believe that there will be a sea battle tomorrow because somebody came back from the future and told me so. Closer to home, the content of a belief may depend on the future: I stipulate that “Lucky” refers to the next winner of the lottery; which truth I believe when I believe that Lucky will win depends on who will win (Hawthorne and Lasonen-Aarnio, 2009, 96). In both cases I may believe an unsettled truth \( p \), and yet not be in danger of forming a false belief that \( p \) at close worlds \( w^* \) at which \( p \) is false, either because I do not acquire the belief in \( w^* \), or because the relevantly similar belief I acquire in \( w^* \) has a different content. Leaving backwards causation aside, unsettled truths are unknowable except under a future-dependent guise.

Last truths thus turn out to be unknowable, except under a future-dependent guise. But they are also much less numerous than one may think. For instance, the proposition that there will never be any centaurs may prima facie seem to be unsettled for about as long as the universe lasts. But if anybody knows it to be true, then it is already settled. (By the same reasoning, if some law is now known it is now settled. But if it is settled, then no world that is close now or later contains a miracle that violates it. Yet the small-miracle conception seems unavoidable in deterministic settings, as we saw in D.1. That puts a strong pressure on the safety theorist to endorse scepticism about the laws. I do not explore the problem further here.)

First truths remain problematic. Let it be a first truth that the value of a certain physical constant \( c \) is \( v \), and suppose that I believe it. If we

---

5. It is worth stressing here that indeterminacy need not imply unsettledness (see sec. D.1 above). Thus the safety condition does not straightforwardly imply scepticism about the indeterminate future. Compare Sturgeon (1993, 160–161).

6. Backwards causation would require us to revise the characterisation of branching-time closeness (C). In the example, there is a close world in which the sea battle will not occur, the time traveller did not tell me that it will, and I did not form the belief. That would be a close world in which the past differs from the actual past, in contradiction with (C).
consider all the histories that were open at the beginning of the universe (or at some arbitrarily early time), we are likely to find some world in which I form a relevantly similar but mistaken belief that \( c = v + 1 \), for instance. So taking the relevant time to be the earliest (or any arbitrarily early time, where there is no earliest) would seem to lead to excessive scepticism. But taking some other time like the time of belief acquisition would seem ad hoc. We leave first truths aside here, and provisionally take the last unsettled time to be the relevant one for other propositions. (As we will see, the proposal runs into difficulties with non-atomic propositions.)

### D.3 Time-relative safety and deductive closure

Let us make three simplifying assumptions. First, we focus on synchronous deduction, that is, deduction that does not take time (Williamson, 2000, 26). Second, we recast any multi-premise deduction from \( p_1, \ldots, p_n \) to \( q \) as a conjunctive deduction from \( p_1, \ldots, p_n \) to \( p_1 \& \ldots \& p_n \) followed by a single-premise deduction from \( p_1 \& \ldots \& p_n \) to \( q \). We thus have only two cases to consider: single-premise deduction and conjunctive deduction.

Third, we assume that:

**Deductive bases (DB).** If at \( w, t \) one believes \( q \) on the basis of competent deduction from premises \( p_1, \ldots, p_n \), then at any \( w^* \) close to \( w \) at \( t_q \), one has a relevantly similar belief that \( q^* \) only if one has respectively relevantly similar beliefs in \( p_1^*, \ldots, p_n^* \) such that \( p_1^*, \ldots, p_n^* \) jointly entail \( q^* \).

That is, relevantly similar conclusion-beliefs are beliefs entailed by relevantly similar premise-beliefs. Safety theorists who uphold closure need to assume (DB). For suppose I competently deduced \( q \) from \( p \) but (DB) fails. Then there is a close world \( w^* \) in which I have a relevantly similar belief that \( q^* \) where \( q^* \) does not follow from beliefs relevantly similar to my actual beliefs in \( p_1, \ldots, p_n \). Then the safety of the later does not

---

guarantee that my belief $q^*$ in $w^*$ is true, so my actual conclusion-belief may be unsafe. For instance, if “competent deduction” is conceived as a fallible process so that some inferential mistakes count as close instances of “competent deduction”, both (DB) and closure fail (Lasonen-Aarnio, 2008, 159-160, 167).

Safety is genuinely time-relative for a subject at $t, w$ iff for some $p, q$ the subject believes at $t, t_p \neq t_q$ and $F(w, t_p) \neq F(w, t_q)$. We show that both in the single-premise case and the conjunctive case, genuine time-relativity is incompatible with either of the following or both:

**Safety closure (SC).** At any $t, w$, if one believes $q$ on the basis of competent deduction from one’s safely true beliefs that $p_1, \ldots, p_n$, then one’s belief that $q$ is safely true.

**Safety counter-closure (SCC).** At any $t, w$, if one believes $q$ on the basis of competent deduction from premises $p_1, \ldots, p_n$, each of which is essential to the deduction, and one’s belief that $q$ is safely true, then one’s beliefs that $p_1, \ldots, p_n$ are safely true.

If (SC) fails and safety is necessary for knowledge, epistemic closure fails.\(^8\) If (SCC) fails, epistemic counter-closure will be violated as well, unless some other condition on knowledge prevents from to gain knowledge by deduction from unknown premises.

Epistemic counter-closure is fairly intuitive: if one does not know that $p$, then deducing $q$ from $p$ cannot generate knowledge that $q$. One may think of counterexamples, though: maybe I can know that a door is more than 1 meter high by deduction from the slightly inaccurate (and therefore false) belief that it is 2 meters high. But in the cases I will discuss, epistemic counter-closure is clearly not violated. So unless some further condition on knowledge is brought up to prevent the conclusion from being known, the cases show the corresponding safety condition to be too weak. As such a condition does not seem forthcoming, and setting inaccuracy cases aside, I will simply assume here that a safety condition

---

\(^8\) Revision after viva. Unless some other condition on knowledge prevents one to know premise beliefs of inferences whose conclusion belief is unsafe.
that violates counter-closure is too weak.\(^9\)

Now, consider single-premise deduction. Suppose that at \(w, t\) one believes \(p\) and competently deduces \(q\). By (DB), at each world in \(F(w, t_p)\) in which one has a belief in some \(q^*\) relevantly similar to one’s actual belief that \(q\), one has a belief in some \(p^*\) that is relevantly similar to one’s actual belief that \(p\), and \(p^*\) entails \(q^*\). By (BS2), one’s actual belief that \(p\) is safe iff one’s relevantly similar beliefs at worlds in \(F(w, t_p)\) are true. The safety of one’s belief that \(p\) thus guarantees the truth of \(q^*\) beliefs at worlds in \(F(w, t_p)\). That in turns entails their truth at worlds in \(F(w, t_q)\) only if \(F(w, t_q)\) is included in \(F(w, t_p)\). Similarly, the unsafety of one’s belief that \(q\) entails that some \(p^*\) belief is mistaken in \(F(w, t_q)\). That in turns entails that one is so mistaken in \(F(w, t_p)\) only if \(F(w, t_p)\) is included in \(F(w, t_q)\). But if safety is genuinely time-relative in the case, \(F(w, t_p) \neq F(w, t_q)\), so at least one of (SC) and (SCC) is not valid, and they are expected to fail in many cases.

Here is one. In May 1999, Alexei learns through the main Russian TV channel that the United States have just bombed a Chinese embassy (\(p\)). The relevant time \(t_p\) is the last unsettled time for \(p\), say April 1999. With appropriate details in place, Alexei’s belief is safely true. In particular, in no world close at April 1999 does Channel 1 falsely claims that the U.S. bombed an embassy. Now Alexei competently deduces that the U.S. have just bombed a Chinese embassy or Alisa has lost a lottery back in 1980 (\(p \lor q\)). (Alexei’s independent grounds for \(q\) are no better than mine in the past lottery case in D.2.) What is the relevant time for \(p \lor q\)? The

\(^9\) The inaccuracy case is based on Williamson’s (2000, 159-160) watermark case; but crucially Williamson’s case does not involve deduction. Luzzi (2011) (to whom I owe the label “counter-closure”) put forwards another purported counterexample: hearing that the TV has been switched on, Ingrid believes that her flatmate is in the lounge (\(p\)). That is true, but only luckily so, for the flatmate is a prankster that often switches the TV on from a distant room. She deduces that the flatmate is in the house (\(q\)). That is true, and safely so, because (unbeknown to her) the flatmate is agoraphobic. On Luzzi’s view, Ingrid’s belief that \(q\) is safely true because “there is no close possible world in which he is not in the house [and] she believes that he is via the method she deploys in the actual world” (2011, 4-5). But a belief relevantly similar to Ingrid’s belief that \(q\) need not be a belief that \(q\). As the case is set up, Ingrid’s belief that the flatmate is in the lounge is relevantly similar to her belief that he is in the house; so the unsafety of the first makes for the unsafety of the second. Pace Luzzi (2011, 7), safety theorists can thus reject the claim that Ingrid knows \(q\).
D.3. Time-relative safety and deductive closure

A proposition was settled back in 1980, when Alisa lost. If we take the last unsettled time to be the relevant one, Alexei’s belief is safe only if there was no danger then that Alexei would form (in 1999) a relevantly similar but mistaken belief. But it may well be that some worlds close then lead to Channel 1’s reports of U.S. bombings being unreliable and Alexei believing them all the same, for it may have been close then that the Soviet Union would last ten years more than it did. Assuming that \( q \) is false in some of these worlds, Alexei’s belief that \( p \lor q \) is not safe, and closure fails. Alternatively, if we take the relevant time to be the last unsettled time among those of the disjuncts, namely 1999, we run into the opposite problem. For suppose now that Alexei deduces \( p \lor q \) from his belief that \( q \). In 1980 there was a danger that Alexei would form (in 1999) a false belief that \( q \), so his belief that \( q \) is not safely true. But in 1999 there is no such danger, as in D.2. Thus there is no danger either that he forms a relevantly similar but mistaken belief that \( p \lor q \). So his belief that \( p \lor q \) is safely true even though his belief that \( q \) is not, and counter-closure fails.

Second, consider conjunctive deduction. Suppose at \( w, t \) one believes \( p_1, \ldots, p_n \) and competently deduces \( q \). By (DB), at each world in \( F(w, t_q) \) in which one has a belief in some \( q^* \) relevantly similar to one’s actual belief that \( q \), one has beliefs in some \( p_1^*, \ldots, p_n^* \) that are relevantly similar to one’s actual beliefs that \( p_1, \ldots, p_n \), respectively. By (BS2), for each \( p_i \), one’s belief \( p_i \) is safe iff one’s relevantly similar beliefs at worlds in \( F(w, t_{p_i}) \) are true. Then the safety of each belief \( p_i \) entails the truth of \( q^* \) beliefs at worlds in \( F(w, t_q) \), which in turns entails their truth at all worlds in \( F(w, t_q) \) only if \( F(w, t_q) \) is included in each \( F(w, t_{p_i}) \). And the unsafety of one’s belief that \( q \) entails that some \( p_i^* \) belief is mistaken in \( F(w, t_q) \), which in turns entails that one is so mistaken in some \( F(w, t_q) \) only if each \( F(w, t_{p_i}) \) is included in \( F(w, t_q) \). But if safety is genuinely time-relative in the case, then either \( F(w, t_q) \neq F(w, t_{p_i}) \) for some \( p_i \), or \( F(w, t_{p_i}) \neq F(w, t_{p_j}) \) for some \( p_i, p_k \), which also ensures that \( F(w, t_q) \neq F(w, t_{p_i}) \) for some \( p_i \). Hence at least one of (SC) or (SCC) is not valid, and they are expected to fail in many cases.
Consider Alexei again, and suppose that he competently deduces $p \& q$ from his beliefs that $p$ and that $q$. His belief that the U.S. have just bombed a Chinese embassy ($p$) is safely true, but his belief that Alisa lost the lottery ($q$) is not. What is the relevant time for $p \& q$? Its last unsettled time is 1999, since $q$ was not settled until then. But at that time, he was not in danger of mistakenly believing that Alisa lost. So if the last unsettled time is the relevant one for the conjunction, Alexei’s belief that $p \& q$ is safe even though his belief that $q$ is not, and counter-closure fails. But if we take the relevant time for a conjunction to be the earliest among the last unsettled time of the conjuncts, we run into the opposite problem. For suppose that Alexei safely believes that Gorbachev was appointed to the Politburo before 1980 ($r$). The last unsettled time for that belief is 1980. Suppose that Alexei competently deduces that $p \& r$. The relevant time for that will be 1980; but in 1980, there was a danger that Alexei would from (in 1999) a mistaken belief that $p$, as before. His belief that $p \& r$ is not safe though both his beliefs in $p$ and $r$ are, and closure fails.

If safety is genuinely time relative, either close or counter-closure fails, both in the single- and multi-premise cases.

**D.4 Generalisation and ways out**

The structure of the problem is easy to see. Call the safety set of a given belief the set of worlds relevant to the safety of that belief. Closure and counter-closure are guaranteed only if the safety set of the conclusion-belief is the same as the safety set of each premise-belief. Since on the time-relative conception (BS2), the safety set of a belief depends on the particular proposition believed, safety sets of premise and conclusion typically differ.

The sensitivity conditions earlier put forward by Nozick (1981, 172–178, 204–211) violated closure for the same reason. On the simplest version, sensitively believing $p$ requires one to avoid false belief that $p$ in all worlds as close as the closest $\neg p$-worlds: what we may call the sensitivity set of a belief thus depends on the particular proposition believed. Since the fact that $p$ entails $q$ does not guarantee that the two corresponding
sensitivity sets coincide, closure fails. (Though interestingly, it guarantees that the set for \( q \) includes that for \( p \); correspondingly, the simple sensitivity condition satisfies counter-closure. However, things are not as straightforward when “methods” (Nozick, 1981, 179) are taken into account.)

Structurally analogous difficulties arise with memory and testimony. Granted, we do not want to claim that any testimony-based belief is knowledge if and only if the testifier knows, nor that a memory-based belief is knowledge if and only if the original belief was. The testifier may be overall unreliable, the hearer excessively gullible, and so on; and similarly for memory. So no straightforward analogue of closure and counter-closure holds here. But (sidestepping some thorny issues) let us say that a testimony or a memory is felicitous when no such problem arises, that is, when the channels of testimonial transmission or memory preservation are all right. A plausible principle is the following:

**Knowledge transfer (KT).** A belief uniquely based on felicitous memory or testimony is knowledge if and only if the original belief was.

In other terms, memory preserves knowledge, but it does not create it; testimony transmits knowledge, but it cannot generate it. (One may learn something from an ignorant testifier, but only with the aid of additional background knowledge.) On a safety view, (KT) requires that safety be closed and counter-closed under felicitous transmission. Correspondingly, if the safety set of a belief depends on which subject holds it, then (KT) is expected to fail for testimony. If the worlds relevant to my knowing \( p \) are not the same as the ones relevant to your knowing \( p \), then my (felicitously) telling you that \( p \) may allow you to know that \( p \) even if I do not, or it may fail to do so even though I do. Analogously, if the safety set of a belief depends on the time at which the subject holds it, then (KT) is expected to fail for memory.

Returning to deductive closure and time-relativity, what options are available to safety theorists? A first one is to reject closure and/or counter-closure. But there are several ways to do this. One can fix a uniform
safety condition on all beliefs. (a) The last unsettled time is the relevant one for all beliefs. Closure fails for single-premise deduction: Alexei may know that the U.S. have bombed an embassy without knowing that they have bombed an embassy or Alisa won a lottery back in 1980 (see D.3). Counter-closure fails for conjunctive deduction: believing without knowing it that Alisa won a lottery back in 1980, Alexei can deduce and come to know that she won it and the U.S. have bombed an embassy. (b) For non-atomic propositions, take the earliest among the last unsettled time of their atoms. Here closure fails for conjunctive deduction, and counter-closure fails for single-premise deduction. Alternatively, one can ensure that one of closure or counter-closure is maintained. (c) Take the lax standard. In the conjunctive case, the last unsettled time is the relevant one; in the single-premise case, the earliest among the last unsettled times of the disjuncts. Closure holds, but counter-closure fails. (d) Take the stricter standard: the last unsettled time in the single-premise case, and the earliest among the last unsettled times of the conjuncts in the conjunctive case. Counter-closure holds, but closure fails.\(^\text{10}\) Violations of counter-closure would be excessively cheap ways of knowing; accordingly, (d) is the less unreasonable path, but it does have to countenance seriously counter-intuitive failures of closure.

Another option is to give up the safety condition on deduced beliefs, and endorse a disjunctive condition: one knows \(p\) only if either one’s belief that \(p\) is deduced from known premises or one’s belief that \(p\) is not deduced and safely true.\(^\text{11}\) That would be particularly disappointing, since safety conditions appeared ideally positioned to explain why deduction is a way to extend one’s knowledge — namely, because it is the safest method of

\(^{10}\) Existential propositions may be considered as disjunctions and universal ones as conjunctions. In non-bergsonian worlds, the conditions (c) and (d) above are not sufficient to guarantee that closure or counter-closure hold, respectively. For instance, the early time at which one conjunct settles may not count as close to the error-possibilities that are relevant to a conjunct settling later, which allows for counter-closure failures even with (d) in place. A solution is to take the safety set of a conclusion-belief to be the intersection of those of its premise-beliefs (lax standard) or their union (stricter standard).

\(^{11}\) See Roush’s (2007) analogous restriction of Nozick’s sensitivity condition to non-deductive beliefs.
belief formation. On the present option, the good epistemic standing of deduction would be an unexplained primitive.

The only option for a full-blown safety theorist who upholds closure and counter-closure is thus to keep safety sets fixed across premise-beliefs and conclusion-belief. There are two ways to do this.

The first is to have a unique safety set per subject and time. (And if we want to avoid structurally similar problems with memory and testimony, we need a single safety set for all subjects and times.) As our Alexei cases suggest (see D.3), it is dubious that we can find such a set by picking up a unique time in a branching-time safety structure. Time-relative safety would probably have to be rejected.

The second is to allow the safety set of a belief to depend on whether it is used in deductive inference. Here again we have the options (a)-(d) above: last unsettled time, earlier time among the last unsettled times of the atoms, lax, and stricter. The difference is that the resulting safety set of the conclusion-belief is now applied to the premise-beliefs as well. Consider option (d), for instance. If one deduces $p \lor q$ from $p$, the relevant time for both premise and conclusion is the last unsettled time for $p \lor q$; that time cannot be later that the last unsettled time for $p$, but it can be earlier. That in turn may “destroy” one’s knowledge that $p$, in the sense that one would have known $p$ if one had not made the deduction. For instance, Alexei would not be in position to know that the U.S. have just bombed an embassy if he deduces from that that they have bombed an embassy or Alicia has won a lottery in 1980. Similarly, if one deduces $p \& q$ from $p$ and $q$, the relevant time is now the earliest one among the last unsettled times of $p$ and $q$, so that one may “lose” knowledge of one premise. For instance, Alexei would not be in position to know that the U.S. have just bombed an embassy if he conjoins that belief with his belief that Gorbachev was appointed to the Politburo before 1980. Option (d) is again the stricter one, in the sense that deduction can only “destroy” knowledge of the premises. Options (a)-(c) would allow deduction to “create” knowledge of the premises. Analogous moves concerning testimony or memory would lead one to views according to which communicating or
remembering one’s beliefs affects their epistemic status.
APPENDIX E

Uncentred possibilities of error

A modal requirement is subject-centred if the errors that it requires to be avoided are only errors involving the same subject. The simple safety requirement is an example of a subject-centred requirement:

**Simple safety** S knows that p only if S could not easily have been mistaken about p.

The requirement says that a subject knows only if it is impossible (in some restricted sense) that the same subject makes a mistake. The vast majority of modal requirements put forward in the literature are subject-centred. Here I will argue that this is a bad idea.

E.1 Problems for subject-centring

Subject-centring may seem natural if we formulate our requirements in terms of avoidance of false belief. There may well be some lunatic around that is prepared to believe that there is fire in the house no matter

---

1. See ch. 3, sec. 3.8.1. Only Armstrong (1973, 168) is very explicit that his account is not subject-centred: “There is some specification of A [the subject] such that, if any person is so specified, if they further believe that p, then p is the case” (emphasis mine). Goldman’s (1976), Lewis’s (1996) and Williamson’s (2000) accounts could be read as uncentred, but it is not entirely clear — see ch.3: 3.4.3, 3.4.6 and the references therein. (Williamson’s (2009a) model is uncentred.) Goldman’s (1979) process reliabilism is uncentred, but it is offered only as an account of justification; to account for knowledge, Goldman (1986, ch.3) uses subject-centring.
what. This does not prevent *me*, who is perfectly sane, from knowing that there is a fire in the house when I see one. The apparent lesson is that only *my* possible errors matter, not those of other people.

However, the lesson is less obvious is safety is relativised to bases or methods, as it usually is (Williamson, 2000, 128; Sosa 1999a, 378; 1999b, 149; 2005, 156; Pritchard, 2005, 152–5):

**Basis-relative safety** \( S \) knows that \( p \) on basis \( b \) only if \( S \) could not easily have been wrong about \( p \) on basis \( b \).

Now it is less obvious that centring is needed. The lunatic would believe that there is a fire even if there was none. But he would not do so on the basis on which I now believe correctly that there is a fire.

Consider the following case:

**Amnesiac Twins** Following a disaster, the twins Matti and Teppo have been separated and have lost memory. Of their past, only a few photos of a younger Matti survive. While each is still ignorant of the existence of his brother, each comes into possession of one of the photos. (We may assume that these are copies of the very same photo.) Seeing the striking resemblance, both Matti and Teppo believe himself to be the boy in the photo.

Matti believes that he [Matti] is the boy on the photo, and he is right. Teppo believes that he [Teppo] is the boy on the photo, and he is wrong. But neither of them knows that they are the boy on the photo.

If Matti is directly referring to the boy that appears on the photo, then the proposition he believes is a necessary truth: namely, that he [Matti] is that boy [Matti]. Safety is trivially satisfied, since there is no possibility in which the proposition is false, and consequently no possibility in which Matti believes it mistakenly. This is, of course, another instance of the necessary truths problem (sec. 4.1). The solution is to give up the centring of safety on a particular proposition:

**Basis safety** \( S \) knows that \( p \) on basis \( b \) only if \( S \) could not easily have been wrong on basis \( b \).

Can Matti’s failure of knowing be explained by this revised safety?
On the basis of seeing a photo of Matti, Matti could not have been mistaken that he is the boy on the photo. *A fortiori*, he could not easily have been mistaken that he is the boy on the photo. So if the basis of Matti’s belief is *seeing a photo of Matti*, then Matti’s belief satisfies basis-relative safety. So safety could not explain why he does not know.

Suppose we rather say that Matti’s basis is that of *seeing a photo of somebody who looks like him*. If Matti saw a photo of Teppo, he would form the false belief that he is the boy on that photo. So if Matti could easily have come into possession of a photo of Teppo, he could easily have formed a false belief on the same basis. Basis-relative safety could explain why he does not know, provided that Matti could easily have seen a photo of Teppo instead. But nothing in the story guarantees that that will be the case. We can alter it so that there could not easily have been photos of Teppo, or that such photos could not easily have reached Matti. For instance, we may suppose that the disaster was precisely the result of the perceived need of someone to destroy all the photos of Teppo. Such variations do not appear to affect the fact that Matti does not know that he is the boy on the photo.

We may ensure that Matti could easily have been mistaken, if we argue that because they are twins, Matti could easily have led Teppo’s life, and conversely. Thus Matti could easily have been in Teppo’s position. That forces a weak reading of “could easily”, but perhaps an acceptable one. (On a time-relative conception of safety, this means that whenever we evaluate a belief of Matti concerning himself, we have to consider all branches of possible history that diverge from actuality from the time of his birth on. See appendix D.)

But consider now the following case:

**The Swampwomen** Walking in a Yorkshire forest, Charlotte and Emily are faced with the choice of crossing a swamp or taking a long detour. They worry whether they may get wet and catch a cold in crossing the swamp, but decide that they will not. Each forms the belief that she will not get sick by crossing the swamp. The swamp is unfortunately infested with ticks that carry a rare
and poorly understood disease that they had never heard of. Charlotte is grounded in bed for the following week, but Emily has an exceptional genetic immunity to it.

We may sharpen the case by assuming that Charlotte and Emily are not parents, and that everybody in Emily’s family has the immunity, though this has not been discovered. There is no sense in which Emily could easily have failed to have the immunity. We may assume that there is no other disease one could get in a swamp in the area; and that Charlotte’s and Emily’s judgement that they could not get a cold was right. Finally, and somewhat artificially, we can make Charlotte and Emily as similar as desired in their internal physical states when they form their belief, but for the difference in genes. Even when the case is so sharpened, it is natural to say that Emily did not know that she would not get sick by crossing the swamp.

With the case so sharpened, Charlotte and Emily form the belief that they will not get sick on the same basis. In Charlotte’s case this leads to a mistake. But in Emily’s case, there is no close possibility in which she forms a false belief on that basis. Basis-relative safety cannot explain why she fails to know.

It may be objected that safety is not meant to explain all failures of knowledge. But it is not clear that it should not explain this one. After all, if Emily had just been temporarily immune to the disease (due to her following a certain treatment), safety would have explained her failure to know. Why should there be an asymmetry between the latter variant and the original version? Moreover, the safety theorist will be hard pressed to say what other requirement may explain Emily’s and Matti’s failure to know. Both are justified, in the common fallibilist sense. Matti has no reason to suspect that someone else on Earth looks exactly like him; Emily has no reason to suspect unknown diseases in the swamp. She can say that Matti’s and Emily’s justification is defeated. But that would be losing one major advantage of safety conditions, namely to replace

---

2. See sec. 4.1.1 for a development of this kind of symmetry argument.
E.2 Uncentred method safety

A straightforward solution to the puzzles is given by giving up subject-centring:

**Uncentred basis safety**  S knows that \( p \) on basis \( b \) only if one could not easily have been mistaken on basis \( b \).

The “one” in the right-hand side is not anaphoric for \( S \). If Teppo forms a false belief on the basis of his looking like the boy in the photo, then one could easily be mistaken on that basis. And if one could easily be mistaken on that basis, Matti cannot come to know on that basis that he is the boy on the photo. Similarly, if Charlotte forms a false belief on the basis of the fact that she will not catch a cold by crossing the swamp, then one could easily be mistaken on that basis, and if so, Emily cannot come to know something on that basis.

What we have called “basis” so far should be properly called method. “Basis” is ambiguous between something like the input of a method and the method itself.\(^3\) If bases are understood as inputs of methods, uncen-tered basis safety makes no sense. Consider all the facts that lead Charlotte and Emily to think that they will not be sick: that their boots are water-proof, that they have but a short walk home afterwards, that they have not got a cold on similar occasions, and so on. Some sufficiently irrational subject could wildly infer about anything from these facts. Suppose a third walker, Anne, infers from these facts that she will never get a cold for the rest of her life, which is false. If the “basis” in Emily’s case is just taken to be the facts in question, then somebody could easily believe something false on that basis: in fact one does, namely Anne. If uncen-tered basis safety is read that way, it implies that Emily and Charlotte do not know that they will not catch a cold; but they do. This is not the way

---

3. See for instance the misunderstanding between Goldman (2009, 84) and Williamson (2009b, 309).
“basis” should be understood for uncentred basis safety to make sense. On the proper understanding, inferring that one will not get a cold ever from those facts is not the same “basis” of belief as inferring that one will not get a cold today from the same facts. It seems to me more natural here to speak of methods: the method that leads one to believe that one will not get cold today on the basis of these facts is not the same as the one that would lead one to believe that one will not ever get a cold from the same facts. So let us say instead:

**Uncentred method safety** \( S \) knows that \( p \) on the basis of method \( m \) only if one could not easily have been mistaken on the basis of \( m \).

Uncentring is not just a trick to deal with a few problematic cases. There is an intuitive sense in which it is just a piece of (good or ill) luck that one is the person one is. And the same holds for many properties one has: there is an intuitive sense in which it is just a piece of luck that one is red-haired rather than blond-haired, born during the day or during the night, and so on. In some cases in which the difference between forming a true belief and a false one merely depends on one of these properties, the resulting true belief is lucky in a way that is incompatible with knowledge. (Not all such properties make for knowledge-preventing luck, of course. I may be lucky enough to have a excellent eyesight; this does not prevent me from acquiring knowledge through it.) When the success of one method of belief of formation depends on it being used by one particular person, then it is a piece of luck that that person gets true belief that way. Subject-centring wrongly classifies beliefs formed that way as safe; the proper evaluation of methods is uncentred.

Here is a case in which ordinary and epistemic sense luck come together:

**The Red-Headed League** The head of the Irish gang is looking for two drivers to help his men to collect payments. Jabez is among the
people approached. Like the others, he is told that the job is safe, and since the boss has a high reputation for keeping his word, he believes it is. He accepts the offer along with another man. Jabez’s job turns out to be safe, but the other driver’s one is dangerous. As it happens, Jabez is red-headed and under no conditions would the boss put a red-headed ordinary person in a dangerous job or lie to them.

It is a piece of luck for Jabez to be red-headed, and it is a piece of luck that the method of trusting the boss is an infallible one for him. The other driver forms a false belief through the same method. On the uncentred safety view, that explains why Jabez does not know: another person could easily have formed a false belief in the same way.

The Red-Headed League is similar to a class of cases in the literature that we may call protective angels cases. These are cases in which some powerful agent ensures that a subject’s belief are not mistaken. Pritchard’s Temperature Room is one (Pritchard, 2010; see p. 213 above); see also Hawthorne (2004, 56n), Roush (2007, 122–3), among others. Protective angels cases are taken to raise a problem for the idea that safety is sufficient for knowledge, since the subject could not easily be wrong in such situations. But given an uncentred version of safety, it is not all clear anymore that such cases pose a threat to the sufficiency of safety. For while there may be an angel that protects me, another subject next to me may use the very same methods as I do and be mistaken. Just as while Jabez may happen to be the boss’s favourite, another person next to him is mistaken in trusting the boss as he does.

E.3 An objection and reply

Uncentred requirements are stronger than centred ones. This raises the worry that they are too strong. A variant of the Swampwomen case brings it forward:

Swamp Travellers Every day, a number of travellers have to cross a
(disease-free) swamp. Most of them form the belief that they will cross it without being sick, as Emily and Charlotte do in the previous case. So does Branwell, but unbeknown to him, he is allergic to a very rare fungus that is found in that swamp. As a result, he is grounded in bed for one week after crossing the swamp.

Uncentring seems to lead to the excessive sceptical conclusion that none of the travellers know that they will cross the swamp without getting sick. For if Branwell’s method for forming his belief is the same as the travellers, then his case shows that one could easily form a false belief on the basis of that method.

One should first note that there is some intuitive pressure to deny knowledge in such cases. This has been well evidenced by Vogel (1990, 20) and Hawthorne (2004, 8) in relation to deductive closure. Unless a particular traveller is accustomed to cross swamps, it is difficult not to accept the following, especially if Branwell’s case is brought to attention:

The traveller does not know that she is not allergic to some rare fungus that is found in this swamp.

But if she does not know that, then it is hard not to accept that:

The traveller does not know that she will cross the swamp without getting sick.

This way of supporting the sceptical intuition may be thought to rely on closure. But the intuition can be pressed directly: once it is pointed out that some apparently healthy person has had an unexpected heart attack,

6. Note that the details are not straightforward. In standard closure-based sceptical arguments of this kind, it is argued that one does not know an ordinary proposition \( p \) that entails a lottery proposition or the negation of a sceptical scenario \( q \). For instance, \( p \) is the proposition that my car is parked at the back of the bar, \( q \) is the proposition that my car is not one of the few cars to have been stolen while their owners where having a drink. The sceptical argument is supposed to use some deductive closure principle to the effect that if one knows \( p \), one is in position to know \( q \). Given the intuitive judgement that one is not in position to know \( q \), we are led to think that one does not know \( p \). But as stated above, the case is not one of entailment. It does not follow from the fact that I have crossed the swamp safely that I was not allergic to some fungus that is found in it. This suggests that closure is either a special case of a more general problem or not at the heart of it at all. On the present view what generates the problematic intuitions is that apparently relevant error possibilities are brought forward.
we easily come to think that we do not know that we will be fine in the coming week.\textsuperscript{7} That is entirely expected on the uncentred view: the error possibilities of other people are directly relevant to one’s knowing. On the centring view, the relevance must be indirect: errors of others are only indicative of errors one could make, and the latter are relevant to whether one knows.

Assuming the sceptical judgement is false, however, how could the uncentred safety view avoid it? There are two options. The first is to limit the sphere of error possibilities that are relevant to a given agent: for instance, only (actual or possible) errors made nearby or at a similar time or within a given community would be relevant to whether some given agent knows. This option replaces subject centring with group or area centring.

The second is to say that the traveller’s method differs from that of Branwell. Plausibly, the difference will not be an internal one. The difference will have to be that the traveller is somehow connected to the facts in a way that is relevantly different from that of Branwell. By basing his beliefs on facts such as whether is shoes are waterproof or whether the outside temperature is too low, the traveller is forming his belief on the basis of facts that are indicative of the fact that she will not get sick. By contrast, when Branwell is forming his belief on the basis of parallel facts, he is not forming his belief on the basis of facts that are indicative of his not getting sick from crossing the swamp. The puzzle for the second view, though, is why the same thing cannot be said about the initial Swampwomen case. Why not say that Emily is also basing her beliefs on facts that are in fact indicative of her not getting sick? To get the issue as precise as possible, compare the original Swampwomen with the following minimal variation:

**Swampwomen-reversed** As in the original case, except that it is Charlotte who has an exceptional sensitivity to the disease. Emily has no special immunity, but the ticks present no special threat to most

\textsuperscript{7} See also the Car theft case, p. 204.
The case is a two-person version of the traveller and Branwell case. Like the traveller, Emily is now an ordinary person who crosses the swamp assuming that it poses no special threat. Charlotte reasons in the same way, but she has an exceptional condition such that the swamp poses a special threat to her.

In the reversed version, it is less clear that Emily does not know. Suppose we want to follow the non-sceptical line on the Traveller and the Reversed Swampwomen by arguing that in each of these cases the error involves a different method. If a rare condition makes for a different method, how can we avoid the conclusion that exceptional genetic immunity makes for a different method in the original Swampwoman case?

Various moves could be explored here; let me just present one. Two preliminaries. First, method safety says that one knows if one bases one’s belief on a method that could not lead to error. It is allowed that a belief is based on several methods. Second, methods are nothing but epistemically relevant classes of belief cases. Conceiving them in this abstract way helps to state the proposed answer.

For many classes of things, there are one or several subclasses of the normal items of the class. (What counts as normal in the class will typically depend on contextual interests; but suppose such a context is fixed.) For instance, a machine produces socks but is not entirely reliable. Some have holes, some are not sewn properly, and so on. We have a subclass of perfectly normal socks: the ones that do not have a defect. Now suppose that in such cases we are allowed only to consider two natural classes: the general class and its normal subclass. Defective socks do not form relevant subclasses of their own.

Applying this picture to our cases, we get the following. In both the Swampwomen and its reversed version, we have two classes: the general traveller class, that includes both subjects, and the normal class, that includes the normal subject only. In the original Swampwomen, Charlotte is the normal subject. She is mistaken, so she does not know.
Emily is the exceptional subject; but she belongs to only one class: the general one. And in this class of beliefs, there are mistakes, namely Charlotte’s. So she does not know either. In the reversed Swampwomen, Charlotte is the exceptional subject. She is mistaken, so she does not know. Emily is the normal subject. She thus belongs to two classes: the general one and the normal one. As a member of the general class of beliefs, her belief is fallibly based, since Charlotte’s belief is in that class and is false. But as a member of the normal class of beliefs, her belief is infallibly based, since no belief in that class is false. So she knows. We thus predict the reversion.  

The gist of the answer is that one is allowed to build finer-grained methods (i.e., classes of beliefs) when these correspond to robust, non-fleeting, non-exceptional features of a subject or situation. Thus we are not allowed to regard Emily’s belief in the original case as belonging to a category of its own in virtue of her exceptional immunity. But we are allowed to regard Emily’s belief in the reversed case in virtue of her not having the exceptional sensitivity.

The outlined answer can prima facie be extended to other cases. In Vogel’s Heart attack and Cart theft cases, the group of healthy persons who know that they will not have a heart attack and the group of people who know where their car is parked can be set apart provided that there is a relevant robust, non-fleeting, non-exceptional feature in their situation that is not present in the exceptional error situations. Not any feature is relevant or acceptable (one cannot group lottery losers as forming a different category), and there is no way to define in advance the relevant or acceptable ones. My interest here is not to defend such an answer about particular cases, but to show that a non-sceptical answer in terms of individuation of methods is available to the defender of uncentred

---

8. To generalise the solution, it is preferable to have many “normality” classes, each resulting from ruling out exceptions of some type: e.g. the class of subjects who do not have exceptional immunity, the class of subjects who do not have exceptional allergy, the class of subjects who have neither. Let me also stress that, despite my formulations, the classes are classes of beliefs, not of subjects. Whether two subjects fall into a same class or not depends on the beliefs under considerations.
safety.

Let me take stock. Subject-centring runs into a range of trouble cases: not only the ones I have put forward, but also the protective angels cases. Uncentring is motivated and explains the trouble cases. Uncentring threatens to lead to excessive scepticism. But we precisely do feel a pressure to accept the sceptical consequences, which indicates that uncentring is not at odds with our conception of knowledge. To avoid the sceptical consequences, a non-sceptical epistemologist can adopt weaker forms of centring, or can individuate methods so as to rule out abnormal errors as relevant to whether one knows in ordinary cases. We have explored one way to do so. Uncentring can be embraced without fear of scepticism.
List of Symbols

Cases and conditions
\(\alpha, \beta, \gamma, \ldots\) variables for cases.
\(C\alpha\) condition: \(C\text{ holds in case } \alpha\).
\(T\alpha\) truth condition: \(T\alpha \text{ iff } p_\alpha \text{ holds in } \alpha\).
\(B\alpha\) belief condition: \(B\alpha \text{ iff in } \alpha, S_\alpha \text{ believes that } p_\alpha \text{ at } t_\alpha\).
\(Bb\alpha\) belief on basis \(b\): \(Bb\alpha \text{ iff in } \alpha, S_\alpha \text{ believes that } p_\alpha \text{ on the basis of } b \text{ at } t_\alpha\).
\(K\alpha\) knowledge condition: \(K\alpha \text{ iff in } \alpha, S_\alpha \text{ knows that } p_\alpha \text{ at } t_\alpha\).
\(E\alpha\) error: \(E\alpha \text{ iff } B\alpha \land \neg T\alpha\).
\(I\alpha\) ignorance: \(I\alpha \text{ iff } T\alpha \land \neg B\alpha\).
\(F\alpha\) falsity: \(F\alpha \text{ iff } \neg T\alpha\).

Methods models

Metalanguage
\(M\) the set of all methods: \(M = W \times (\mathcal{P}(P) \times \mathcal{P}(P))\), where \(W\) is the set of worlds and \(P\) the set of propositions.
\(p, q, \ldots\) propositions
\(\pi, \pi'\) sets of propositions (premises, conclusions)
\(m, n, \ldots\) methods
\(0, 1\) the empty method: \(0(w, \pi) = \emptyset\) for any \(w, \pi\). the identity method: \(1(w, \pi) = \pi\) for any \(w, \pi\).
\( m(w, \pi) \) the set of conclusions reached by method \( m \) at \( w \) from premises \( \pi \)

\( m(w) \) the set of conclusions reached by method \( m \) at \( w \) unconditionally: \( m(w) = m(w, \emptyset) \).

\( m + n \) the union of methods \( m \) and \( n \): for any \( w, \pi \), \( (m + n)(w, \pi) = m(w, \pi) \cup n(w, \pi) \).

\( m \circ n \) the combination of \( m \) and \( n \): for any \( w, \pi \), \( (m \circ n)(w, \pi) = m(w, n(w, \pi)) \).

\( \mathfrak{F} = \langle W, M^B, R \rangle \) a methods frame. \( M^B \subseteq M \) is the set of basis methods of the agent. \( R \subseteq W \times W \) is a reflexive accessibility relation over worlds.

\( \mathfrak{M} = \langle \mathfrak{F}, V \rangle \) a methods model. \( \mathfrak{F} \) is a methods frame and \( V \) is a valuation that maps propositional constants and methods constants to propositions and methods, respectively.

\( M \) the methods set of the agent. In our models, \( M \) is the closure of \( M^B \) under composition and union.

\( M^I(w) \) the set of infallible methods at \( w \). For any \( m, w, m \in M^I(w) \) iff \( \forall w', wRw': \forall p (p \in m(w') \rightarrow w' \in p) \).

\( B(m, w) \) the set of propositions believed on the basis of \( m \) at \( w \). For any \( m, w, B(m, w) = m(w) \) if \( m \in M \), and = \( \emptyset \) otherwise.

\( K(m, w) \) the set of propositions known on the basis of \( m \) at \( w \). For any \( m, w, K(m, w) = B(m, w) \) if \( m \in M^I(w) \), and = \( \emptyset \) otherwise.

\( B(w) \) the set of propositions believed at \( w \). \( B(w) = \bigcup_{m \in M} B(m, w) \).

\( K(w) \) the set of propositions known at \( w \). \( K(w) = \bigcup_{m \in M} K(m, w) \).

\( m^R \) the Pure Reason method. \( m^R(w, \pi) = \{W\} \) for any \( w, \pi \).

\( m^D \) the Deduction method. \( m^D(w, \pi) = \{p : \exists q, r \in \pi (q \cap r \subseteq p)\} \) for any \( w, \pi \).

\( m^{SD} \) the Single-Premise Deduction method. \( m^{SD}(w, \pi) = \{p : \exists q \in \pi (q \subseteq p)\} \) for any \( w, \pi \).
\( m^{\text{CD}} \) the Conjunctive Deduction method. \( m^{\text{CD}}(w, \pi) = \{ p : \exists q, r \in \pi(p = q \cap r) \} \) for any \( w, \pi \).

\( b_p \) the proposition that \( p \) is believed: \( b_p := \{ w : p \in B(w) \} \).

\( -b_p \) the proposition that \( p \) is not believed: \( -b_p := \{ w : p \notin B(w) \} = W \setminus b_p \).

\( k_p \) the proposition that \( p \) is known: \( k_p := \{ w : p \in K(w) \} \).

\( -k_p \) the proposition that \( p \) is not known: \( -k_p := \{ w : p \notin K(w) \} = W \setminus k_p \).

\( m^{\text{pi}} \) a positive introspection method: a method \( m^{\text{pi}} \in M \) such that \( m^{\text{pi}}(w, \pi) = \{ b_p : p \in \pi \} \) for any \( w, \pi \).

\( m^{\text{ni}} \) a negative introspection method: a method \( m^{\text{ni}} \in M \) such that \( m^{\text{ni}}(w, \pi) = \{ -b_p : p \notin B(w) \} \) for any \( w, \pi \).

\( m^{\text{pc}} \) a positive confident introspection method: a method \( m^{\text{pc}} \in M \) such that \( m^{\text{pc}}(w, \pi) = \{ k_p : p \in \pi \} \) for any \( w, \pi \).

\( m^{\text{nc}} \) a negative introspection method: a method \( m^{\text{nc}} \in M \) such that \( m^{\text{nc}}(w, \pi) = \{ -k_p : p \notin B(w) \} \) for any \( w, \pi \).

\( \mathcal{F} = \langle W, N \rangle \) a neighbourhood frame. \( N \subseteq W \times \mathcal{P}(\mathcal{P}(W)) \) is a neighbourhood function that maps worlds to sets of propositions (sets of sets of worlds).

\( \mathcal{M} = \langle \mathcal{F}, V \rangle \) a neighbourhood model. \( V \) is a valuation function that maps propositional constants to sets of worlds.

\( I(m, w) \) the doxastic information given by method \( m \) at \( w \): \( I(m, w) = \bigcap B(m, w) \).

\( E(m, w) \) the epistemic information given by method \( m \) at \( w \): \( E(m, w) = \bigcap K(m, w) \).

\( I(w) \) the agent’s doxastic information: \( I(w) = \bigcap B(w) \).

\( E(w) \) the agent’s epistemic information: \( E(w) = \bigcap K(w) \).

\( R^{\text{im}} \) doxastic accessibility relation for method \( m \): \( wR^{\text{im}}w' \) iff \( w' \in I(m, w) \).
List of Symbols

$R^E_m$ epistemic accessibility relation for method $m$: $wR^E_m w'$ iff $w' \in E(m, w)$.

$R^I$ doxastic accessibility relation: $wR^I w'$ iff $w' \in I(w)$.

$R^E$ epistemic accessibility relation: $wR^E w'$ iff $w' \in E(w)$.

Object language

$m, n, \ldots$ constants for methods.

$m^R, m^D, m^{CD}, m^{SD}$ constants for the Pure Reason, Deduction, Conjunctive Deduction, Single-Premise Deduction methods. For any model $\mathcal{M}$, $[m^R] = m^R$, $[m^D] = m^D$, $[m^{CD}] = m^{CD}$, $[m^{SD}] = m^{SD}$.

$m^{pi}, m^{ni}, m^{pc}, m^{nc}$ constants for a positive introspection, negative introspection, positive confident introspection, and a negative confident introspection methods. For a given model $\mathcal{M}$ and some $m^{pi}, m^{ni}, m^{pc}, m^{nc}$, $[m^{pi}] = m^{pi}$, $[m^{ni}] = m^{ni}$, $[m^{pc}] = m^{pc}$, $[m^{nc}] = m^{nc}$.

$\mu, \nu, \ldots$ variables for methods.

$\mu + \nu, \mu \circ \nu$ the union of $\mu$ and $\nu$, the composition of $\mu$ and $\nu$.

$Bm : p$ the agent believes $p$ on the basis of $m$.

$Km : p$ the agent knows $p$ on the basis of $m$.

$Bp$ the agent believes $p$.

$Kp$ the agent knows $p$.

$\square p$ necessarily, $p$.

Semantics

$[\cdot]_{\mathcal{M}}$ The interpretation function for a model $\mathcal{M}$.

$\models_{\mathcal{M}} \phi$ $\phi$ is true in model $\mathcal{M}$ at $w$. 
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Typology of requirements</td>
<td>109</td>
</tr>
<tr>
<td>3.2</td>
<td>Simple Infallibility Requirements</td>
<td>147</td>
</tr>
<tr>
<td>3.3</td>
<td>Parametric infallibility and reliability requirements</td>
<td>148</td>
</tr>
<tr>
<td>3.4</td>
<td>Double requirements</td>
<td>149</td>
</tr>
<tr>
<td>7.1</td>
<td>Four classes of modal auxiliaries</td>
<td>333</td>
</tr>
</tbody>
</table>
List of Figures

4.1 Lottery cases and Gettier cases ......................... 179
4.2 The Gettier problem for reliability requirements .......... 180
4.3 Method infallibilism ........................................ 181

5.1 The prime number case ........................................ 239
5.2 The fake orange case ........................................ 240
5.3 Chisholm’s Sheep case ....................................... 241
5.4 The original Gettier case .................................... 242
5.5 Inductive knowledge ........................................... 244

C.1 Counterexample to $M_B$ and $M_K$. ........................ 420
C.2 Counterexample to $N_K$. ..................................... 421
C.3 Counterexample to $K_K$. .................................... 422
C.4 Counterexample to $4_K$. .................................... 422
C.5 Informative introspection. .................................... 430


Bibliography


Ichikawa, J. (ms.). Knowledge and counterfactuals. aug. 2008.


ion to Locke’s “Essay concerning human understanding”*, pages 313–351. 
Cambridge University Press.

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*.

Cambridge, Mass.

Locke’s “Essay concerning human understanding”*, pages 406–435. Cam-
bridge University Press.

pages 1688–1702.

Pappas, G. S. and Swain, M. (1973). Some conclusive reasons and ’con-

Pasnau, R. (2010). Medieval social epistemology: *Scientia* for mere mort-


Brace.

497.

Press.


57(2):140–145.


Williamson, T. (2009c). Reply to john hawthorne and maria lasonen-
aarnio. In Greenough, P. and Pritchard, D., editors, *Williamson on


University Library.


71:429–472.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Aris-

tivity and surreptitiously raising standards. *The Philosophical Quarterly*,

This page intentionally left blank.
Detailed Contents

Short Contents iii
Acknowledgements vii
Abstract xiii
Résumé xvii

1 Introduction 1
1.1 Method infallibilism and conceptual analysis . . . . . . . 1
1.2 What methods are . . . . . . . . . . . . . . . . . . . . . . . . . 4
1.3 Methods infallibilism and the theory of knowledge . . . . 6
1.4 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

I History of Epistemology 13

2 The Legend of the Justified True Belief Analysis 15
2.1 The Legend and its puzzles . . . . . . . . . . . . . . . . . . . . 19
2.1.1 Why is it so hard to find any clear statement of the justified true belief analysis before the XXth century? 20
2.1.2 Why did nobody notice Gettier cases? . . . . . . . . . 24
2.1.3 Why were Gettier-style cases presented as cases of true belief that is not knowledge? . . . . . . . . . . . . . . 25
2.1.4 Why was the justified true belief analysis not introduced as the traditional one? . . . . . . . . . . . . . 27
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.5 How the Legend appeared</td>
<td>30</td>
</tr>
<tr>
<td>2.2 Ur-foundationalism</td>
<td>31</td>
</tr>
<tr>
<td>2.2.1 Marks of truth</td>
<td>32</td>
</tr>
<tr>
<td>2.2.2 Truth preserving inferences</td>
<td>35</td>
</tr>
<tr>
<td>2.2.3 Naive Ur-foundationalism</td>
<td>35</td>
</tr>
<tr>
<td>2.2.4 Discernibility and access</td>
<td>37</td>
</tr>
<tr>
<td>2.2.5 Discernibility and luminosity</td>
<td>39</td>
</tr>
<tr>
<td>2.2.6 Basing</td>
<td>44</td>
</tr>
<tr>
<td>2.2.7 Coherentist Ur-foundationalism</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Historical illustrations</td>
<td>46</td>
</tr>
<tr>
<td>2.3.1 Stoics on basic knowledge</td>
<td>47</td>
</tr>
<tr>
<td>2.3.2 Were Stoics externalists?</td>
<td>53</td>
</tr>
<tr>
<td>2.3.3 Aristotle, Epicurus, Stoics, Descartes</td>
<td>62</td>
</tr>
<tr>
<td>2.3.4 Induction as truth preserving inference</td>
<td>65</td>
</tr>
<tr>
<td>2.4 Hume’s predicament, probabilism and idealism</td>
<td>67</td>
</tr>
<tr>
<td>2.4.1 Sceptical objections to Ur-foundationalism</td>
<td>68</td>
</tr>
<tr>
<td>2.4.2 Ur-foundationalist scepticism</td>
<td>74</td>
</tr>
<tr>
<td>2.4.3 Idealism</td>
<td>81</td>
</tr>
<tr>
<td>2.5 The Gettier problem and the internalism/externalism divide</td>
<td>84</td>
</tr>
<tr>
<td>2.5.1 Fallibilist internalism: Malcolm, Ayer and Chisholm</td>
<td>85</td>
</tr>
<tr>
<td>2.5.2 Philo of Larissa, John Buridan, and Thomas Reid:</td>
<td>88</td>
</tr>
<tr>
<td>2.5.3 The Gettier problem and externalism</td>
<td>90</td>
</tr>
<tr>
<td>2.5.4 A dilemma for mixed internalist views</td>
<td>91</td>
</tr>
<tr>
<td>2.5.5 The contemporary focus on justification</td>
<td>94</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>95</td>
</tr>
</tbody>
</table>

**II Epistemology**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Modal Requirements</td>
<td>101</td>
</tr>
<tr>
<td>3.1 Conditions and requirements</td>
<td>103</td>
</tr>
<tr>
<td>3.1.1 Cases, conditions, and requirements</td>
<td>103</td>
</tr>
<tr>
<td>3.1.2 Fallible and infallible conditions</td>
<td>105</td>
</tr>
</tbody>
</table>
3.1.3 Typology of requirements ............................................. 108
3.2 Motivation for modal requirements ................................. 109
  3.2.1 Infallibilist requirements need not be modal ............... 109
  3.2.2 Methodological considerations for infallibility require-
           ments ................................................................. 112
3.3 The avoidance of error .................................................. 116
  3.3.1 Error, ignorance, falsity .......................................... 116
  3.3.2 Overview of modal requirements .............................. 117
3.4 Simple modal requirements ........................................... 120
  3.4.1 The schema .......................................................... 120
  3.4.2 True belief and infallible belief .............................. 120
  3.4.3 Goldman’s relevant-alternatives condition and Lewis’s
         infallibilist contextualism ...................................... 121
  3.4.4 Nozick’s sensitivity and tracking ............................. 122
  3.4.5 Sosa’s safety and tracking ...................................... 126
  3.4.6 Williamson’s safety .............................................. 132
3.5 Parametric modal requirements ....................................... 133
  3.5.1 The schema .......................................................... 133
  3.5.2 Armstrong’s nomological infallibilism ....................... 135
  3.5.3 Nozick’s sensitivity and tracking with methods .......... 136
  3.5.4 Basis-relative safety ............................................ 138
3.6 Reliability requirements .............................................. 139
  3.6.1 The schema generalised .......................................... 139
  3.6.2 Goldman’s process reliabilism ................................. 140
  3.6.3 Pritchard’s most-worlds safety ............................... 142
3.7 Double modal requirements ........................................... 143
  3.7.1 The schema .......................................................... 143
  3.7.2 Goldman’s global and local reliability ..................... 143
  3.7.3 Pritchard’s double safety account ............................ 144
  3.7.4 Sosa’s and safety ............................................... 144
3.8 Discussion of the results ............................................. 146
  3.8.1 Proposition- and subject- centring ........................... 150
  3.8.2 Variability .......................................................... 151
4 Methods infallibilism

4.1 The Method constraint
155
  4.1.1 The necessary truths problem 156
  4.1.2 Method safety 164
  4.1.3 Relevantly similar cases safety 166
  4.1.4 Conclusion 169

4.2 The Infallibility constraint
171
  4.2.1 The fallibilist dilemma 172
  4.2.2 The structure in ordinary cases 174
  4.2.3 The Gettier problem for fallibilism 175
  4.2.4 Formal version of the argument 177
  4.2.5 Conclusion 182

4.3 Objections to necessity
183
  4.3.1 The generality problem 183
  4.3.2 The concept of knowledge is primitive 186
  4.3.3 Methods can be more basic than knowledge 189
  4.3.4 Application: fine-grained, non-luminous methods 192
  4.3.5 Scepticism and possibility 200
  4.3.6 Conclusion 206

4.4 Objections to sufficiency
207
  4.4.1 Internal justification 208
  4.4.2 The Absence of Defeaters 211
  4.4.3 Reliability and virtue 213
  4.4.4 Simplicity 218

4.5 Conclusion 221

III Formal Epistemology

5 Methods-based Models
225
  5.1 Introduction 226
  5.2 Methods 231
    5.2.1 The space of possible methods 231
    5.2.2 Operations on methods 233
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.3</td>
<td>Infallibility</td>
<td>235</td>
</tr>
<tr>
<td>5.3</td>
<td>Applications</td>
<td>238</td>
</tr>
<tr>
<td>5.3.1</td>
<td>The prime number case and fake-barn-style Gettier cases</td>
<td>239</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Standard Gettier cases</td>
<td>241</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Fallible but reliable methods</td>
<td>242</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Inductive knowledge</td>
<td>243</td>
</tr>
<tr>
<td>5.4</td>
<td>Methods models</td>
<td>245</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Propositions and the problem of Frege cases</td>
<td>245</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Frames and agents</td>
<td>248</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Language</td>
<td>251</td>
</tr>
<tr>
<td>5.5</td>
<td>Results</td>
<td>253</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Subjectivity, factivity, and referential transparency</td>
<td>254</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Perfect reasoning</td>
<td>257</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Consistency</td>
<td>261</td>
</tr>
<tr>
<td>5.5.4</td>
<td>Perfect introspection and perfect confidence</td>
<td>262</td>
</tr>
<tr>
<td>5.5.5</td>
<td>Excellence</td>
<td>267</td>
</tr>
<tr>
<td>5.5.6</td>
<td>Discussion</td>
<td>269</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusion</td>
<td>270</td>
</tr>
<tr>
<td>IV</td>
<td>Semantics</td>
<td>273</td>
</tr>
<tr>
<td>6</td>
<td>Modal Epistemic Contextualism</td>
<td>275</td>
</tr>
<tr>
<td>6.1</td>
<td>Epistemic contextualism</td>
<td>277</td>
</tr>
<tr>
<td>6.2</td>
<td>Modal Epistemic Contextualism: Lewis</td>
<td>280</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Outline</td>
<td>280</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Subject-sensitive vs. context-sensitive rules</td>
<td>282</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Lewis’s rules are specific to “know”</td>
<td>284</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Possibilities and attending to possibilities</td>
<td>285</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Formal version of Lewis’s view</td>
<td>288</td>
</tr>
<tr>
<td>6.3</td>
<td>Modal Epistemic Contextualism: DeRose</td>
<td>289</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Outline</td>
<td>289</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Closeness and methods</td>
<td>290</td>
</tr>
</tbody>
</table>
6.3.3 The Rule of Sensitivity ........................................ 292
6.3.4 DeRose vs. Lewis on context shifts ....................... 293
6.3.5 The Rule of Sensitivity is specific to know, and has
no clear analogue elsewhere in language .................. 294
6.3.6 Formal version of DeRose’s view ......................... 298
6.4 The necessary truths problems (again) ..................... 299
6.4.1 Lewis’s “linguistic disguise” suggestion ................ 299
6.4.2 The proper basing answer .................................. 301
6.4.3 Methods contextualism ..................................... 304
6.5 Attending, presupposing and taking seriously ............ 306
6.5.1 Attending is not sufficient to take seriously .......... 306
6.5.2 The no-belief account of taking seriously ............ 307
6.5.3 The problem of third-person sceptical arguments ... 309
6.6 The unmentioned subjects problem ......................... 312
6.6.1 Semantics for unuttered sentences ..................... 312
6.6.2 Factivity fails in Lewis’s semantics ..................... 313
6.6.3 Diagnosis: Lewisean vs. DeRose’s semantics .......... 315
6.6.4 A puzzle for DeRose’s Rule of Sensitivity .......... 317
6.7 Conclusion ...................................................... 320

7 Modals and knowledge ........................................... 323
7.1 The semantics of modals ...................................... 324
7.1.1 Kratzer’s semantics for modals ......................... 324
7.1.2 Hacquard on root vs. high modals ..................... 330
7.2 Circumstantial modals and real possibility ............... 336
7.2.1 Disagreement cases ....................................... 337
7.2.2 Conjunction cases ....................................... 340
7.2.3 The proposal ............................................... 341
7.2.4 Caveats .................................................... 342
7.3 Subjunctive conditionals and real possibility ............ 345
7.3.1 von Fintel and Gillies on subjunctive conditionals ... 346
7.3.2 Unified real possibility semantics ...................... 348
7.3.3 Real and theoretical possibilities ....................... 349
### Detailed Contents

7.4 The real possibility account of knowledge attributions . . . 351  
7.4.1 Semantics of “Know” ........................................ 352  
7.4.2 Taking a possibility seriously ................................. 353  
7.4.3 Comparison with Lewis’s and DeRose’s accounts ........ 355  
7.4.4 Knowledge ascriptions and subjunctive conditionals 359  
7.4.5 Conclusion ...................................................... 362  

7.5 Contextualism and invariantism ................................. 362  
7.5.1 Naturalness ...................................................... 363  
7.5.2 Semantic blindness ............................................ 365  
7.5.3 Knowledge and Action ....................................... 369  
7.5.4 The price of invariantism ................................... 370  

8 Conclusion ................................................................... 373  
8.1 History ................................................................. 374  
8.2 Epistemology ........................................................ 375  
8.3 Formal epistemology .............................................. 378  
8.4 Semantics ............................................................. 380  
8.5 The two sources of scepticism .................................. 382  

V Appendices .................................................................. 385  

A Subjunctive Conditionals and Tracking .......................... 387  
A.1 Semantics for subjunctive conditionals ....................... 387  
A.2 Variably strict semantics ....................................... 391  
A.2.1 Lewis’s semantics ............................................. 391  
A.2.2 Lewis-Nozick semantics .................................... 391  
A.2.3 Quasi-Nozickian semantics ................................. 392  
A.2.4 Nozick’s Stretch semantics ................................. 393  
A.2.5 Strict Spread semantics ..................................... 395  
A.2.6 Nozick’s Weakened Spread semantics ................. 396  
A.3 Context-sensitively strict semantics ......................... 398  
A.4 Hybrid semantics ................................................ 400  
A.5 Nozick’s tracking .................................................. 402