

Synthese Library 382

Studies in Epistemology, Logic, Methodology,  
and Philosophy of Science

Jody Azzouni

# The Rule- Following Paradox and its Implications for Metaphysics



Springer

# Synthese Library

Studies in Epistemology, Logic, Methodology,  
and Philosophy of Science

Volume 382

## **Editor-in-Chief**

Otávio Bueno, University of Miami, Department of Philosophy, USA

## **Editorial Board**

Berit Brogaard, University of Miami, USA

Anjan Chakravartty, University of Notre Dame, USA

Steven French, University of Leeds, UK

Catarina Dutilh Novaes, University of Groningen, The Netherlands

The aim of *Synthese Library* is to provide a forum for the best current work in the methodology and philosophy of science and in epistemology. A wide variety of different approaches have traditionally been represented in the Library, and every effort is made to maintain this variety, not for its own sake, but because we believe that there are many fruitful and illuminating approaches to the philosophy of science and related disciplines.

Special attention is paid to methodological studies which illustrate the interplay of empirical and philosophical viewpoints and to contributions to the formal (logical, set-theoretical, mathematical, information-theoretical, decision-theoretical, etc.) methodology of empirical sciences. Likewise, the applications of logical methods to epistemology as well as philosophically and methodologically relevant studies in logic are strongly encouraged. The emphasis on logic will be tempered by interest in the psychological, historical, and sociological aspects of science.

Besides monographs *Synthese Library* publishes thematically unified anthologies and edited volumes with a well-defined topical focus inside the aim and scope of the book series. The contributions in the volumes are expected to be focused and structurally organized in accordance with the central theme(s), and should be tied together by an extensive editorial introduction or set of introductions if the volume is divided into parts. An extensive bibliography and index are mandatory.

More information about this series at <http://www.springer.com/series/6607>

Jody Azzouni

# The Rule-Following Paradox and its Implications for Metaphysics

 Springer

Jody Azzouni  
Department of Philosophy  
Tufts University  
Medford, MA, USA

Synthese Library  
ISBN 978-3-319-49060-1 ISBN 978-3-319-49061-8 (eBook)  
DOI 10.1007/978-3-319-49061-8

Library of Congress Control Number: 2017930681

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Acknowledgments

This book is essentially the doubling in size of an earlier paper on the same topic: “The rule-following paradox and the impossibility of private rule-following” (*The Baltic International Yearbook of Cognition, Logic and Communication*, Volume 5: *Meaning, understanding and knowledge*, 2010). My gratitude similarly doubles the acknowledgments of help received for the original article and help received for the subsequent book. So here goes. I must thank, to begin with, Brendon Larver and David Corfield for organizing the conference *Two streams in the philosophy of mathematics: Rival conceptions of mathematical proof* (July 1–3, 2009, University of Hertfordshire, Hatfield, UK) and, of course, for inviting me to speak. (I may have disappointed them. Only when I arrived did I realize they wanted me to speak about the Azzouni-Rav debate—which I’d never *heard of* until then—but I had prepared to speak about this instead.) I also thank the scientific committee, and the invited organizers, of the 5th International Symposium of Cognition, Logic and Communication for the conference *Meaning, understanding and knowledge* (August 7–9, 2009, Riga, Latvia) and, again, for inviting me to speak. I especially want to thank Jurgis Šķilters, Anda Baklane, and the participants of both conferences for making them such rich and enjoyable experiences.

I thank Douglas Patterson for his detailed comments on the penultimate draft of my original article. They were extremely valuable, as always. My thanks as well to Eric Schliesser for directing me to some invaluable terminology and for always making wise suggestions about the many ways I might explore diplomacy in print.

I should add that this book (as well as the paper it originates from) discharges an old promise I made in *Metaphysical Myths, Mathematical Practice: The Ontology and Epistemology of the Exact Sciences* (Cambridge, 1994). I indicated there that my philosophical approach to mathematical practice has, as a corollary, a response to (a version of) the rule-following paradox. In spirit, although not in expository details, the original article is what I had in mind. The characterization of the evolution of scientific language that I wrote up in 1994–1995 and subsequently published in *Knowledge and Reference in Empirical Science* (Routledge, 2000)—especially Part IV—illustrates the “waxing and waning picture” of language, as well as the contrastivist/noncontrastivist characterizations of language evolution; but reasons

of space prevented an explicit application of these notions to the rule-following paradox at that time. I'm thus very grateful to the editors of the *5th issue of The Baltic International Yearbook of Cognition, Logic and Communication* for giving me the opportunity to finally lay all this out.

I taught the penultimate version of this book in the fall of 2015, and I received invaluable suggestions and “pushback”—much of which I dealt with in the final rewriting of the book in the summer of 2016. My thanks to Brad Clendenen, Matthew DiRe, Megan Entwistle, Casey Lange, Jeff McConnell, Steven Norris, Taylor Oddleifson, George Smith, Michael Veldman, and Douglas Yetman. My thanks to Tufts University and its philosophy department for creating an environment that makes possible research courses like this one. I also thank the anonymous referee for several suggestions about exposition, organization, and content that I've taken. I'm grateful to both Otávio Bueno and Ties Nijssen for smoothing this book's road to publication.

Samia Hesni should be thanked for taking me (“dragging me,” is perhaps more accurate) to the Pacific APA 2016 symposium on the normativity of meaning. I was only glancingly aware of much of that literature—and it gave me an opportunity to do a fresh run-through of a number of articles that raise concerns that, I'll admit, I'm largely unsympathetic to. Nevertheless, some commentary is required, which I added to the manuscript during its final revision. My thanks to Eric Dean for some editorial suggestions on the final version.

I'm grateful to Mario De Caro for looking over the nearly final version of the manuscript in December of 2016 and pressing me (in various ways) to engineer additional clarity. Apart from that, I also want to thank him for urging me, more than once, to expand my original article on rule following into a book. He gave good reasons, but I resisted for years, as I always resist suggestions like this. I have so many other projects I'm hoping to finish (before I die or otherwise enjoy some permanent form of cognitive impairment). Two subsequent events, however, changed my mind. First, I ran across an article of J.R.G. Williams, published in 2007; and I realized that his defense of reference magnetism deserved a response in this context. Second, Otávio Bueno urged me to participate in the Brief Books series of Springer; and this meant that a large book was ruled out from the start. This is important, because I need external constraints like word limits in order to, well, limit how much I tell the reader at one go. The original article covers so much territory precisely because the editors insisted there were no word limits—despite my desperate pleas for them. Even though this book ended up in a different Springer series, it's still the shortest philosophy book I've ever written. The efficacy of word limits, everyone should have them—everyone like *me*, anyway.

# Contents

<b>1</b>	<b>General Introduction .....</b>	<b>1</b>
<b>2</b>	<b>Kripke's Version of Wittgenstein's Paradox and His Solution .....</b>	<b>9</b>
2.1	Setting Up the Problem: Three Constraints on Any Solution.....	9
2.2	Why Introspective Resources Are Insufficient for a Response to Meaning Scepticism.....	15
2.3	How Kripke's Three Constraints Block Dispositional Approaches.....	19
2.4	An Attempted Defense of Dispositional Approaches to Meaning Scepticism.....	22
2.5	The Failure of a Straight Sociological Solution to the Rule- Following Paradox .....	25
2.6	Assertability Conditions, Not Truth Conditions.....	28
<b>3</b>	<b>Two Versions of Robinson Crusoe .....</b>	<b>33</b>
3.1	A Community of Idiolects.....	33
3.2	Robinson Crusoe in Empirically Favorable Circumstances.....	37
3.3	A Robinson Crusoe with Two Sets of Dispositions That He Has Conscious Access to.....	41
3.4	What Does "Better Matching the World" Mean to Crusoe 2? .....	43
3.5	Introspection.....	48
3.6	Last Remarks.....	53
<b>4</b>	<b>Reference Magnetism .....</b>	<b>55</b>
4.1	A Second Family of Straight Solutions to the Rule- Following Problem: Reference-Magnetism Approaches .....	55
4.2	Can Crusoe Use the Way Things Are in the World as Standards for His Words? .....	56
4.3	A Top-Down Approach to Reference-Magnetism: Williams's Interpretationalism .....	61



4.4	Moorean Considerations Plus the Only Game in Town Argument.....	66
4.5	Concluding Remarks.....	72
<b>5</b>	<b>How Positive Success Curves Enable Private Rule Following .....</b>	<b>73</b>
5.1	A Robinson Crusoe with Continuously Varying Dispositions that He Has Conscious Access to.....	73
5.2	A Robinson Crusoe with Plpci Dispositions That He Has—at Best—Inferential Access to.....	78
5.3	Being Guided by Rules Rather Than Compelled by Them.....	83
5.4	Is It Misleading to Describe the Standards for Crusoe 5's Words as "External Ones"? .....	84
<b>6</b>	<b>Truth and Falsity Attributions and Truth-Conditional Semantics in Private Languages.....</b>	<b>87</b>
6.1	How Crusoe 5 Speaks of the Truth and Falsity of His Own Statements and How He Talks About What His Words Refer to .....	87
6.2	Truth Conditions, Bivalence, and the Broad Ignorance Thesis—Applied to Our Own Language.....	91
6.3	Tarski Biconditionals and Truth-Conditional Semantics—Applied to Our Own Language.....	96
6.4	Tarski Biconditionals and Truth-Conditional Semantics—Applied to Crusoe 5's Language.....	102
6.5	Some Concluding Remarks.....	105
<b>7</b>	<b>Correspondence Metaphysics and the Cogency of a God's Eye View .....</b>	<b>109</b>
7.1	Introduction .....	109
7.2	Contrastivist and Noncontrastivist Views of Private Languages.....	110
7.3	Is Our Concept of a God's Eye View Coherent? .....	116
7.4	Epistemic Modesty.....	119
	<b>References .....</b>	<b>121</b>

# Chapter 1

## General Introduction

**Abstract** A summary of the entire book is given. Apart from descriptions of the contents of each chapter, several additional methodological points are made. I give reasons to avoid, as tools of philosophical analysis, concepts such as “understanding,” “meaning,” and “fact.” I also describe some significant differences between how I understand rule following and how Kripke does. In particular (and this is a difference between my approach and that of most philosophers concerned with this topic), I focus very much on “rule following” as it occurs in the application of tasks that the subject engages in during his interaction with the world, as opposed to intrinsic arithmetic exercises, such as counting numerals or adding them.

Central to the “rule-following paradox,” which is the topic of this book, is the notion of “understanding,” as used in “she understands how to add,” “he understands how to continue counting numbers above ten,” and “she understands those concepts.” A methodological warning is immediately in order, however: “Understanding,” so used, is a word for a bewilderingly intricate human capacity that involves subpersonal, conscious/phenomenological, sociological, and normative elements. It’s a notion that’s too dangerously complex to be *presupposed* in philosophical analyses; it should only be a *target* of such analyses. It’s hardly the only word too dangerous for philosophers to use as a tool for analysis rather than as a target of it. Another over-used serpentine trap is “explanation”; yet a third is “meaning.” Even repeated and obvious (and palpable) failure doesn’t seem to stop philosophers from continuing to build fine-tuned analyses on such foundations of sand.

Quine rarely comments on his general approach to doing philosophy—even his avowals of naturalism and extensionalism (for example) are pretty specific in content and application. Nevertheless, I’ve found a rare methodological meditation that’s worth *memorizing*. Quine (1981, 184) writes:

“Sentences have replaced thoughts,” according to Schulenfrei’s account of my views, “and dispositions to assent have replaced belief.” Does he mean that for me there is no more than this to thought and belief? Reading on, I suspect that he does. Then he misunderstands me.

Quine continues:

My position is that the notions of thought and belief are very worthy objects of philosophical and scientific clarification and analysis, and that they are in equal measure very ill suited

for use as instruments of philosophical and scientific clarification and analysis. If some one accepts these notions outright for such use, I am at a loss to imagine what he can have deemed more in need of clarification and analysis than the things he has thus accepted.

But let's stick with the notion of "understanding" since that's the notion this book is directly concerned with. Common to several philosophical traditions is the assumption that someone's understanding of a set of concepts must be grounded in the grasping of the rules that govern those concepts. This captures—so it's thought—the way we understand the meanings of words, as well as how we understand the concepts involved in simple mathematical practices—such as counting. A nice set of models for this picture can be found in rule-governed games, such as chess. This line of thought predicts that once the rules of chess are learned by an individual, she will understand how to play chess—any chess game under any set of circumstances.

Some proponents of this position describe us as having "dispositions" to grasp those rules, or even as having such rules embodied in our dispositions so that we behave appropriately in situations where we exhibit this kind of "understanding"—for example, when counting the oranges in a basket.

A huge literature has focused on "the rule-following paradox"—that aspects of our rule-following practices seem at odds with this natural picture of our grasp of rules. In particular, Kripke (1982) has spawned a large philosophical industry focused on the problems he presents Wittgenstein as having raised for any approach that grounds the understanding of simple mathematical rules in the exercise of a set of dispositions. (Kripke does not present the view as his own; and he—and others—recognize it may not be Wittgenstein's. The coined word "Kripkenstein," consequently, has entered the literature. I avoid this term in this book, preferring to let phrases like "Kripke's Wittgenstein," or context, do its work instead.)

Briefly (and perhaps inaccurately), the rule-following puzzle is this: We understand ourselves as following rules, and given appropriate training, as able to do this. But the means by which we understand ourselves as abled in this way are shown by Kripke as not up to the job. The candidates for enabling the rule-following ability, for example, an introspective grasp of rules, various habits or dispositions to execute rules when faced with a task, and so on, are shown by Kripke as unable to underwrite all the required elements of our competence for rule following. Kripke responds to the paradox by offering a "sceptical solution": Such a solution, conceding failure to the "meaning sceptic," reconfigures the understanding of our practice of rule following in several crucial ways. Among them is that an understanding of meaning in terms of "truth conditions" is rejected. Instead, what's required are "assertability conditions" in terms of the conformity of an individual's rule-following practices with that of a community.

Kripke's version of Wittgenstein's rule-following paradox has been extremely influential. One of my concerns is with how it—and Wittgenstein's views more generally—have been perceived as undercutting the individualistic picture of mathematical practice: the view that individuals (I'll call them "*Robinson Crusoes*"), independently of a community, can engage in cogent mathematics, and indeed can

have “private languages.” The paradox denies that phrases like “correctly counting” can be applied to such individuals because these normative notions can only be cogently applied relative to community standards (so the analysis is taken to reveal).

One of the original elements in my alternative “sceptical solution” to the rule following paradox is that this shocking corollary doesn’t follow—not even if Kripke’s Wittgensteinian objections to dispositional approaches to rule following are largely right. This is because my solution to the rule-following paradox doesn’t favor community standards over individual ones. Furthermore, it doesn’t replace truth conditions with assertability conditions; and this is essential to Kripke’s solution favoring the community over the individual. Instead, my approach focuses on how dispositions to execute rules enable individuals to interact more or less successfully with the world. Suppose their practices exhibit dispositions that change in a way that over time progressively optimizes success events (e.g., predicting how much food will be needed over how many days). Then conformity between how individuals follow rules will result even without rule-following practices being explicitly related to a community.<sup>1</sup>

Lewis (1983), in an article almost as influential as Kripke (1982), connects Kripke’s version of the rule-following paradox to Putnam’s (1981) then version of anti-realism by virtue of offering the same solution to both of them. Lewis urges an approach that, nearly enough, builds into the references of natural-kind terms (as well as mathematical terms such as “sum”) a *presupposition* of metaphysical constraints on the possible extensions of those terms. Lewis contends that this presupposition is *required* to respond to Kripke’s Wittgensteinian paradox *and* to Putnam’s anti-realism. A subsequent generation of analytic metaphysicians has explicitly adopted Lewis’s response under the rubric of “reference magnetism.” In Chap. 4, I show that this family of Lewisian approaches to the rule-following problem—hereafter described as “reference-magnetism” approaches, following the nomenclature of Lewis’s followers—doesn’t work and isn’t needed.

Lewis never uses “reference magnetism.” This phrase apparently originates in Hodes (1984, 135), where he expresses a puzzle. Hodes’s question is: how is it that,

we all end up speaking languages in which, out of all possible numberers [that is, type 2 functions  $F$  carrying type 1 concepts to objects, where for all such concepts  $X$  and  $Y$ ,  $F(X) = F(Y) \equiv (Q_{Ex})(Xx, Yx)$ ], the phrase “the number of” stands for the standard numberer? Why is the standard numberer a “reference magnet” which “draws” reference by that phrase, in a way in which its nonstandard competitors cannot?

“Reference magnetism,” is well-tuned (as Hodes’s phrases so often are); no wonder it was adopted by the next generation of philosophers to convey something different.

Here’s an outline of the chapters of the book.

In Chap. 2, I revisit Kripke’s discussion of the rule-following paradox for dispositional approaches to numerical competence, and the alternative solution to the paradox that he offers on behalf of Wittgenstein. To a large extent, the considerations

---

<sup>1</sup>I’m omitting important details about how this works out in order to give a clear overview. The subsequent chapters don’t omit these details.

I raise in this chapter are Kripke's original ones. In doing so, I don't worry about questions of Wittgensteinian exegesis—of how true Kripke's concerns are to Wittgenstein's originals.<sup>2</sup> Kripke's formulation of the challenges to the meaning-dispositionalist is significant (and extremely influential) regardless of how true it is to Wittgenstein's own thinking. There are two changes, however. First, Kripke's discussion uses the central example of a subject adding, and the sceptical challenge that everything she has learnt and previously experienced is consistent with the possibility that she isn't adding but instead doing something different. I substitute for addition the more elementary task of counting, and the corresponding possibility of doing something different instead of counting. This is minor. Kripke (1982, 17) himself raises the issue about counting in passing.

The second change is significant: it's the key to my solution to the rule-following paradox. My central rule-following paradox-case focuses on the subject who is counting *things*—objects—not someone who is solving a task that involves numerals *alone*. Children learn to count in just this way: they learn to *apply* counting-numbers to tasks of recognizing cardinal numbers of sets of objects, a kind of knowledge that takes several years to acquire.<sup>3</sup> This is important because “social solutions” to the rule-following paradox—ones that normatively favor the community over the individual—have been seen by many philosophers (and sociologists of knowledge) to be the only successful responses that can be made to the paradox. One reason for this is a neglect of considerations about mathematical-concept application *to the world*. Similarly, application of mathematics to the world is largely off-stage in Kripke's discussion (and, as a knock-on effect, it's largely off-stage in the discussions of the many commentators on this literature).

It's not particularly off-stage in Wittgenstein's own work (it's important to add) if only because he often gives examples of people counting objects. Kripke also takes note of the numerical sizes of collections often being the target of counting tasks when he discusses—in passing—“quounting,” and when he discusses Wittgenstein's text. However: if the counting of collections of *objects* is centrally incorporated into the examples that Kripke's meaning sceptic challenges, as I show in later chapters, that makes salient possible (but overlooked) coherent forms of private-language practices. In particular, the crucial notion of *private-language-practice coherence-inducing dispositions* (*plpci dispositions*) that I introduce in Chap. 5 has *no* foothold outside contexts of applying concepts to the worldly items they hold of. My sceptical solution to the rule-following paradox, therefore, won't help disembodied Cartesian entities that avoid boredom by (eternally) counting numerals in their heads.

These disagreements with Kripke's Wittgenstein aside, I should add that I'm in partial agreement with one important lemma that Kripke draws from the rule-following paradox, and that he (1982, 78–79), in a rare moment, officially endorses:

---

<sup>2</sup>There is literature on this question. A taste: Blackburn (1984), Goldfarb (1985, 1992), McDowell (1984), Tait (1986).

<sup>3</sup>See, for good discussions of this, and the citation of relevant literature, Carey (2009) and Butterworth (1999), chapter 3, sections 1–4. A classic study is Gelman and Gallistel (1986).

One must reject “the natural presupposition that meaningful declarative sentences must purport to correspond to facts.” (He adds: “The picture of correspondence-to-facts must be cleared away before we can begin with the sceptical problem.”) I’m in *partial* agreement with Kripke because I think that rule-following considerations do show that we must accept that *some* meaningful declarative sentences don’t correspond to facts. I don’t, however, think that this is established with respect to *all* meaningful declarative sentences. I’ll show why rule-following considerations demand only a *partial* rejection of correspondence metaphysics in Chap. 6. This discussion simultaneously motivates my particular form of truth deflationism—one that I’ve extensively discussed elsewhere.<sup>4</sup>

One crucial point to make now is this. The dispositionalist hope that so much of Kripke’s (1982) analysis is dedicated to crushing locates the *grounds*, of the correspondence relation between meaningful true sentences and the facts to which they correspond, within the “mind” (broadly described) of the person following a rule. There are facts *about the person* and what she *can do* that determine the correspondence relations between her meaningful true sentences and facts in the world. I think rule-following considerations show this dispositional project fails; dispositions *cannot* do the job that most dispositionalists have required of them, and this is largely for the reasons that Kripke has given (on behalf of Wittgenstein).

I reject, however, a major thesis that many draw from this lemma. Kripke has carefully described himself as only an expositor, but a large number of philosophers (and sociologists of knowledge) assert the conclusion that considerations, like the ones I’ll present in Chap. 2, lead inexorably to the result that standards for mathematical practice—and for rule following in general—must be relativized to, or embodied in, the *community* within which the individual learns rules. To use Kripke’s (1982, 109) language,

What is really denied is what might be called the ‘private model’ of rule following, that the notion of a person following a given rule is to be analyzed simply in terms of facts about the rule follower and the rule follower alone, without reference to his membership in a wider community.

One aim of this book is to show that although Kripke is largely right about the nature of the rule-following paradox and the responses following from it, the private model of rule following emerges unscathed from the paradox he presents. The contours of logical space in this problem-area are more twisted and subtle than late last-century thinkers anticipated.

---

<sup>4</sup>See Azzouni (2006, forthcoming), and Chap. 6 of this book. Interpreting Kripke’s formulation, above, is complicated because of his phrase “must purport.” My view is that meaningful declarative sentences can or even “must” purport to correspond to the facts; that’s compatible with some of them *not* actually so corresponding. (The cogency of this gloss depends, of course, on exactly what “purport” means. The word isn’t entirely clear. Also, whether Kripke’s “clearing away” is intended to be a complete or only partial removal of “correspondence” isn’t clear either.) See the discussion on the use of the knowledge/ignorance idioms in Sect. 6.2, and specifically footnote 15 of Chap. 6, for my ways of maneuvering all this.

After I sketch Kripke's exposition of a particular argumentative strategy for his solution to the paradox—his interpretation of Wittgenstein as replacing truth conditions with assertability conditions—I'll argue (in subsequent chapters) for another solution; one that doesn't favor community standards over individual standards. In this chapter, I also briefly indicate why a dispositional approach that broadens the location of the needed dispositions to those possessed by the individuals in an entire community doesn't succeed.

In Chap. 3, I explore disposition-meaning languages: languages that really are intrinsically private. Speakers of these languages (I call them "Crusoes"), unlike us, really do intend each word to refer to whatever their dispositions incline them to apply that word to. I begin the exploration of the scope of such languages, showing how certain notions we commonly use: "mistake," "justification," and "concept," for example, either survive only in a diminished form or vanish altogether. In either case, very little damage to private rule following results. My aim—in this chapter (and in Chap. 5)—is to show exactly how these notions break down or aren't needed in the context of disposition-meaning languages. I also begin an exploration of the content of phrases like "fitting the world," or "carving the world at its joints" for speakers of such languages. One unexpected corollary is that, contrary to the impression Kripke's discussion gives, a community of speakers of such languages can successfully use these languages to engage with the world under a broad range of favorable empirical circumstances. Furthermore, under such circumstances, speakers can objectively compare different disposition-meaning languages, and determine which are better for interacting with the world. They will *not*, however, use notions such as one language "better fitting" the world than another, nor notions like "the correct" application of a word. The phrase, "favorable empirical circumstances," I should add, labels more than characterizations of external-world circumstances: it also covers empirical facts about the mechanisms of the dispositions *in* such speakers—in particular, mechanisms that determine how their dispositions can *evolve*.<sup>5</sup>

In Chap. 4, as I indicated before, I need to pause in the analysis of disposition-meaning languages in order to deflect reference-magnetism solutions to the challenges I've posed so far for such languages to determine references to collections of things in the world. I discuss three versions of the reference-magnetism approach. The first treats the world's structure as, somehow, metaphysically providing resources that supplement what individuals in a community bring to determine reference, so that the result tames the rule-following paradox. Our words (concepts) have determinate reference beyond the psychological and neurophysiological resources of any individual, and beyond what any community of such individuals

---

<sup>5</sup>Here's a cute corollary of my solution to the rule-following paradox that I won't describe any further outside this footnote. An issue arises for the view that artificial languages are various useful tools that one can adopt or drop, depending on one's purposes—a view found originally in Carnap (1956). How is it possible to compare the virtues and vices of such languages *cogently* without, as a result, ascending to a metalanguage of some sort from the vantage point of which these specific languages can be compared? The solution to this puzzle arises right here.



would be able to manage on its own. The second treats interpreters of natural languages as required by semantic theory (along with background scientific practice) to impose determinate reference on the terms of those languages, along with natural kinds as the relata of the kind terms of those languages. The third treats the *a priori* constitutive imposition of natural kinds to be required by the Moorean facts of determinate reference and by semantic theories (that presuppose determinate reference) being “the only game in town.” I show that none of these defenses of reference magnetism work.

In Chap. 5, I present my final Crusoe (Crusoe 5), who is psychologically close to us in the sense that he doesn’t naturally treat his own language *as* a disposition-meaning language because (like us) his dispositions to use terms aren’t visible to him. I show how, nevertheless, his awareness of how his applications of terms can improve his overall well-being over time, enables him to impose coherent standards on his private-language practice.

I examine Crusoe 5’s language more closely in Chap. 6 to show how he can apply a standard truth-functional semantics to it, and how he can employ the truth idiom naturally, similar to the ways we do. This will show that Kripke’s substitution of assertability conditions for truth conditions in order to provide a “sceptical solution” to the rule-following paradox isn’t the only option: another sceptical solution is possible. The required substitution of assertability conditions for truth conditions, however, is essential to the conclusion that private languages are impossible.

In Chap. 7, I attend to some remaining loose ends. I explore the methodological role of a God’s eye view in my approach to the evaluation of private rule followers. To this end, a distinction is drawn between two possible views of Crusoe’s languages (and our languages), one that treats such languages as ever-changing with regard to what their terms refer to, and a second that takes reference as fixed and unchanging. I explore and contrast these two viewpoints to show what kind of sense, if any, we can make of a God’s eye view.<sup>6</sup>

Moving on, I indicate the consequences for correspondence metaphysics that *do* follow from my sceptical solution to the rule-following paradox. Quickly put, the possibility of an empirically-sustained correspondence metaphysics is *not* ruled out. What *is* ruled out is a presuppositional role for such metaphysics in philosophical explanations of empirical success, or in semantics.

Finally, I conceptually connect the rule-following paradox to Hume’s problem of induction.

---

<sup>6</sup>I first offered these different ways of seeing the operation of language in Azzouni (2000), Part IV.



## Chapter 2

# Kripke's Version of Wittgenstein's Paradox and His Solution

**Abstract** This chapter reviews Kripke's original description of Wittgenstein's paradox and its solution. Quite a bit of critical commentary has engaged with Kripke's interpretation; but I'm not concerned with whether Kripke's interpretation of Wittgenstein is right; I'm only concerned with Kripke's puzzle, as he presents it. Two distinctions discussed in the chapter are Kripke's description of straight and sceptical solutions, and that between grounding facts and correspondence facts. Three key discussions occur in the chapter. First, there is a discussion of Kripke's three conditions on a straight solution to the paradox, the infinitude requirement, the justification requirement, and the mistake requirement. Second, there is a discussion of why dispositional approaches to the paradox don't work. Third, there is a discussion of why a straight sociological solution doesn't work. I engage as well with some of the critical commentary on rule following that has arisen subsequently to Kripke's book.

### 2.1 Setting Up the Problem: Three Constraints on Any Solution

We may describe a subject who has successfully counted various collections of different kinds of objects as having learned to count correctly. Then we take it that she will go on "in the same way" to count in new cases: larger numbers of objects, perhaps, than she has counted in the past, as well as the numbers of collections of other kinds of objects (pears instead of oranges, widgets instead of fruit, collections containing assorted heterogeneous items ...).<sup>1</sup> Let's say that she has been taught a numeration system that provides a rule for generating new numerals from earlier ones. Systems of numerals are strikingly different in this respect from the sets of

---

<sup>1</sup> Some of the tacit knowledge required to successfully do this is labeled, by Gelman and Gallistel (1986, 77–82), as the counting principles: "cardinal word," "order-irrelevance," and "abstractness"—that the last word in a count is the cardinal number of the collection, that the order in which the objects in a collection are counted doesn't matter, and that it doesn't matter what the objects are. These abilities are acquired by children in stages from ages 2 to 4. See Carey (2009, p. 241–244) for a description of the process, and for indications of how arduous acquisition of tacit counting knowledge is.

verbal names for numbers that occur in most ordinary languages.<sup>2</sup> The latter usually belong to a finitary notation for numbers that requires the explicit coining of new vocabulary for ever-larger numbers, as in English and other natural languages. The notational generation of ever more numerals, however, is built into numeral systems—subject (of course) to the physical constraints of the medium of the numerals. Once a subject has acquired this ability to count, and has mastered a particular numeration system, we describe her as *understanding* how to count.<sup>3</sup>

Now the subject attempts to count the items in a collection that's larger than any she has ever counted before: it has 57 items. She gets an answer, the correct answer we'd say. Kripke's sceptic challenges the claim that our subject's current answer is in accord with her previous understanding of her task. In the past, Kripke's meaning sceptic claims, she never meant to count, she always meant to quount.<sup>4</sup>

---

<sup>2</sup> See Butterworth (1999), especially p. 52–62, for an accessible discussion—with some citations—of the different types of number vocabulary to be found in different natural languages.

<sup>3</sup> Children can recognize the indefinite nature of numbers without having learned a genuine numeral system. Carey (2009, 252) quotes one five-year old as saying: “suppose you think a gazillion is the highest number—well, you can go a gazillion and one, a gazillion and two ...,” and she describes the spontaneous invention of arguments that there is no highest number as common. In my exposition, I assume the subject has acquired a numeral system for counting in order to avoid complications with number-languages that are outstripped—at least in this respect—by the subject's knowledge of number itself. In any case, children in our culture do, in time, acquire an understanding of numeral systems. That achievement, too, is arduous because of confusions that can arise from the child needing to identify the terms of a numeral system with the number-words in natural languages. Apart from this there is empirical evidence that different brain circuits handle these different number systems. (See, for example, the numerous results on this by Dehaene and his associates. Some of this empirical evidence turns on the loss of certain abilities while others are retained—after strokes and other sorts of specifically-localized brain injuries; but many of them, especially recently, turn on brain imaging during the execution of various arithmetic tasks.)

<sup>4</sup> Kripke (1982, 16): “Quounting” a heap of objects is counting in the ordinary sense unless the heap is formed as the union of two heaps, one of which has 57 or more items (assuming the subject has never so-far counted heaps that large), in which case the answer is 5. Also, see Gelman and Gallistel (1986, 51), where an interestingly similar thought experiment is raised.

Kripke's strategy of undermining what a subject means to do at a present moment by first undermining what the subject could have meant on previous occasions has been widely criticized and (in my view) often misunderstood. Forbes (1983–1984, 225–226) interprets Kripke's “claim that there is no fact about what the subject meant in the past [as ambiguous] between the claim that he had no determinate understanding in the past and the claim that although he had such an understanding, there is no fact about whether it is the same as his present one. The context dictates that it is the second reading which is intended ...” No, the context indicates that the first reading is intended. The challenge to the subject, by the sceptic, that perhaps she meant quounting and not counting, undercuts possible crosstemporal identity relations between present understanding and past understanding by undercutting the capacity of any past facts about the subject to make determinate whether she *meant* counting or quounting. The incapacity of present facts about the subject to make determinate whether she means counting or quounting then follows because present facts about the subject are no different in their capacity to fix meaning than past facts are. On the other hand, Boghossian (1989, 515) interprets Kripke's puzzle as *directly* about what, if anything, determines content- or meaning-conditions—“the [possession] of a correctness condition” instead of Kripke's puzzle only being indirectly about this via the antecedent question of what a subject could have meant on previous occasions. This is a straightforward misreading of Kripke's text. As Forbes

To counter this sceptical claim, that the subject intended to quount, not count, and therefore to show that her current answer really does conform with her previous understanding of what she was doing when she undertook the task in the past, what's needed are the facts about the subject—or the pattern of facts—that underwrite her previous and current understanding of counting. I'll introduce the phrase “grounding facts” to describe the facts—for example, ones about the dispositions of the subject—that explain what the understanding of that subject comes to.

*Some Terminology* I distinguish “correspondence facts” from “grounding facts.” The former are the purported (external-world) facts that true meaningful sentences correspond to. The latter are the psychological/dispositional facts in the person (or elsewhere in the environment) that underwrite a subject's capacity to understand true meaningful sentences, and consequently that determine that such sentences are indeed understood to correspond to the (external-world) facts that they purportedly correspond to. As we'll see shortly, the meaning sceptic's challenge directly places pressure on the purported grounding facts; this in turn undermines correspondence facts.

“Fact” is like those treacherous words I began the general introduction worrying about. I intend nothing philosophically heavyweight by my use of it. In particular, there are no entities—*facts*—that sentences correspond to (I presume there is *nothing*—metaphysically speaking—that sentences correspond to). Some sentences have terms that refer, and how it is with what those terms refer to are “facts.” (See, however, Chap. 6 for refinements of this idea to handle the common cases where sentences contain terms that don't refer.)<sup>5</sup> I also don't mean to allude to a metaphysically heavyweight notion of “grounding,” or to any of the complex literature that's recently bloomed on this topic.<sup>6</sup> “Fact” and “grounding”—as in “grounding fact”—are ordinary pre-theoretical terms.

Before resuming the dialectic, I should describe a second bit of terminology. Kripke distinguishes between “sceptical” and “straight solutions” to the rule-following paradox. Straight solutions seek out grounding facts (of one sort or another) for the subject's apparent understanding of counting. Sceptical solutions accept that grounding facts don't exist, and explain the subject's grasp of counting in some other way.

*Back to the Dialectic* As noted, it's the purported grounding facts (or the pattern of grounding facts) that account for it being true that in the past the subject understood how to count, rather than it being true that she understood (and intended) to “quount.” This relevant pattern of grounding facts is something *about* the subject, either something about her mental states, or something about the subpersonal events or structures that underlie those mental states that enabled her to grasp counting. The

---

(1984, 226) correctly notes, Wright's (1980) concern (not Kripke's) is directly about what, if anything, determines content- or meaning-conditions. (Forbes regards Wright's claim as “more direct and more challenging” than Kripke's; I think he thinks this because of his misreading of Kripke's strategy that I attribute to him at the beginning of this paragraph.)

<sup>5</sup>Also see Azzouni (2012c).

<sup>6</sup>As discussed in, e.g., Correia and Schnieder (2012).

purported grounding facts about the subject's understanding are supposed to be explained in *these* terms: how the subject's psychological states—what the subject thought when facing the task—enabled her to understand herself to be counting, or how mechanisms or subpersonal representations (neurophysiological or otherwise) activate so as to have enabled the subject to be able to count.<sup>7</sup>

I read Kripke as placing three constraints on the required account of the pattern of grounding facts, dispositional or otherwise, that are supposed to constitute a subject's understanding of counting (and that are supposed to provide a response to the meaning sceptic).<sup>8</sup> *The infinitude requirement*. Any given subject has counted only finitely many collections of objects. Her presumed ability to count the many collections of objects that she has never counted before raises fresh considerations in two ways. First, as noted earlier, she may have only counted apples and pears, but nothing else. Second, she may have counted collections of no more than 57 objects. We take her grasp of counting to include an ability to count collections of objects that differ—both in number and in kind—from collections she has counted in the past. We think that something is wrong if a subject hasn't realized that counting is neutral about the objects she can count—if she feels unable to count red objects, or ones kept in certain boxes, and so on. We also feel something is wrong if how she counts larger collections of objects deviates from how she counts smaller collections of objects. Perhaps she systematically (or asystematically) skips numerals, starts over when she reaches a certain numeral, or fixates at a particular numeral as the answer, and tells us that the rest of the objects aren't relevant.<sup>9</sup> Part of what we assume "understanding" to mean is that the subject can handle counting tasks that are different from ones she has handled in the past.

Kripke's second constraint on any account of the grounding facts about the subject that explains how she understands counting is *the justification requirement*.

---

<sup>7</sup>Carey (2009), for example, offers an "ontogenetic" description of how the child—already possessing certain innate subpersonal cognitive systems (the parallel individuation of small sets, and the natural-language quantifiers) applicable to particular (restricted) numerical tasks—is enabled over the course of a year and a half by induction and analogy ("Quinean bootstrapping") to understand certain crucial properties of numerals, e.g., that they continue indefinitely, and to understand how to apply them to count collections of objects. A striking effect of the rule-following paradox, as we'll see, is that Carey's empirical hypothesis, and similar others on offer by cognitive scientists, don't even look *relevant* to a solution.

<sup>8</sup>Ginsborg (2011, 228) describes three objections she (and others) take Kripke to raise against the dispositional account; these resemble but differ in significant respects from the requirements I read him as placing on grounding facts. See what follows and Sect. 2.3.

<sup>9</sup>Some of these possible "mistakes" actually occur during a child's acquisition of counting but most don't. Children go through specific stages as they acquire counting skills, and cognitive scientists describe children successively, as "one-knowers," "two-knowers," "three-knowers," "four-knowers," "subset-knowers," and finally, "cardinal-principle knowers." They are described—at one stage—as "one-knowers" because they know how to distinguish one object from many, but they can't distinguish between the cardinal numbers of groups of objects larger than one; "two-knowers," are analogous, and so on. At one stage in their acquisition of counting skills, children also skip numerals when counting; they don't do this later. See Carey (2009) and Butterworth (1999) for descriptions, and for references to the literature.

When the subject counts a collection of objects, and recognizes the number of objects present, her answer is “justified.” She *ought* to have proceeded as she did given what she *meant*. It’s no accident that her way of counting gives the expected answer.

Equally crucially, adults and children take themselves to be justified in the answers they give *because* they take themselves to have grasped how to count: they take themselves to understand what “counting” *means*. If someone is asked *why* she counts a certain way, her response will be: “Because *that’s* how you count.” Or: “Because *that’s* how it’s done.” Or: “Because *that’s* what counting *is*.” If a child (or adult) counts in an unusual way, she’ll justify herself by showing (or indicating that she believes) that the unusual method always results in the same answers as strict enumerative counting does.<sup>10</sup>

Notice that the justification offered—*that’s* how it’s done, etc.—is, except in unusual cases, phenomenologically confident. The subject, once she has learned the procedure, doesn’t hesitate in how she goes about it. And, she doesn’t hesitate in giving an explanation for what she does: “This is what counting *is*.” Also, we accept this justification from a third-person point of view when we say: She *does* understand what “counting” means. (Or: She *does* understand how to count.)

There is a subtlety here. As I described “justification,” we normally have *two* expectations about a child (or anyone) with respect to counting—and with respect to other tasks as well. The first is that, given the child *means* to count, what she does conforms with her intention. Sometimes we fail at a task precisely because what we do *doesn’t* conform to what we *meant* to do; we don’t do what we “ought” to have done. (“Look what you’ve done! Did you really mean to do *that*?”) The second expectation is that what the child (or anyone) means to do is *count*. A different kind of mistake can occur—the person can mistake the task at hand, think she is “supposed” to do something else. Here the required “ought phrase” is different too; it’s not that “she ought to have proceeded in the way she did given what she meant,” but rather that “she meant to do task T but she ought to have meant to do task U.” The phenomenological confidence that’s felt when undertaking an assigned task involves both expectations and can correspondingly be undermined in two ways.

There is a bit of literature about the supposed “normativity” involved in rule following—specifically, how thick the “ought” is.<sup>11</sup> I interpret this “ought” as thin.

<sup>10</sup>“First I group them into collections of five because they’re easier to recognize. I count all the groups, and then multiply by five.” This is a description that a child might give because it strikes her as obvious that her new method gives the same answers as the first method she was taught (strict enumeration).

<sup>11</sup>That is, there is an on-going debate about “content normativity” or “the normativity of meaning”—in what sense, if any, normativity is involved in rule following or meaning determination. Some such view about the normativity of meaning is attributed to Kripke on the basis of his being taken to claim that “one is or was *disposed* to respond in a certain way on a given occasion cannot make it the case that one *ought* so to respond” (as Ginsborg (2011, 228) puts it, *italics hers*). See, for example, Boghossian (1989, 2003), Glüer (1999), Wikforss (2001). As I indicate above, and later, “normativity,” “rationality,” “norms,” “correctness” or “appropriateness,” are involved in this topic only in the sense that these *presuppose* successful rule following, and *not* in the sense that

"Justification," here, is meant only in the sense that given what the child *meant*, what the child does *conforms with* this meaning. The justification requirement, that is, presupposes that "built into" what the child means (or meant) is enough to support that her execution conforms to what's meant. This is what's behind the phenomenological confidence of a child—that she can "explain" why what she's doing is what she means to do—what she intends to do—and that she can explain that what she means to do is what's being asked of her. Both the richness of what's meant (that *it* determines what behavior conforms to it) and *our* grasp of that meaning and what it implies are at work here.

As I've indicated in the preceding paragraphs, the justification requirement is closely linked to a third requirement Kripke gives, *the mistake requirement*. Confidence in how we count, and in the results we get when we count, is open to a particular "rider" that everyone recognizes. This is that a "mistake" hasn't been made. Any subject can make mistakes when counting a collection of objects, and get wrong answers—this is routine. We clearly and naturally distinguish cases where a subject has made a mistake from cases where we describe the subject as not having (fully) grasped the concept of counting (and, notably, we also distinguish cases where a subject has made a mistake—even a systematic mistake—from cases where: (i) we would describe the subject as not having yet acquired the concept of counting, and from cases where: (ii) we would instead say something seems to be "wrong" *with the subject*.

Overlooking an object, or counting one twice (by accident—because of distraction, say) are cases we describe as ones where a subject "has made a mistake." But something seems more seriously wrong if the subject always deliberately counts all the red objects twice over, to get double the sizes of those collections. If the subject isn't just "playing games," but thinks she's counting, we're moved to deny she understands counting. Under certain circumstances, we may say that she's incapable of learning *how* to count. An important aspect of our identifying the subject as having made a mistake (although this isn't a necessary condition) is that the subject can be brought to *recognize* that she has made a mistake. ("You skipped that one." "Oops," she says, correcting herself.)

However we describe the subject's ability to count (as embodied in her) and however we characterize the disposition she has acquired by learning to count, room must be made for a concomitant ability to get *wrong* answers, and to *recognize* that they *are* wrong answers. We don't normally speak of an "ability" to get wrong answers; but the requirement should be clear nevertheless. We often get wrong answers even though we grasp counting perfectly well, and any description of the

---

any of these are presupposed by rule following. Part of this point is that the "ought" currently in play is unremarkably hypothetical, only of the form (put somewhat archly), "If you'd like to live, you *ought not* to jump off a bridge." ("If you'd like to live, I'd recommend ....") Similarly, "if you really want to live, it isn't *appropriate* to jump off a bridge, now is it?" (Imagine this is said, in a movie, by an angel to a potential suicide.) Gibbard (2003, 85) writes: "... whether we ought to take a walk can depend on the weather; that doesn't make the weather normative in any philosophically special sense." So, to this extent, I'm on the side of those in this debate who reject "meaning normativity." I say more about this in Sect. 2.3.

pattern of grounding facts that we use to explain anyone's ability to count—the ability to execute any task, for that matter—has to allow room for the possibility of errors, and for the possibility of recognizing those errors.

I should stress one last related point about how compatible making mistakes—and even systematic *tendencies* to make mistakes—is with our attribution of the understanding of concepts, such as counting, addition, subtraction, multiplication, and division. Some people are amazing calculators, quickly and accurately solving numerical problems involving large numbers. It's not felt that these people *grasp* these concepts in a way that the rest of us don't—more “deeply,” say. It's just that such people can calculate faster, and know more tricks and shortcuts (in many cases, they have simply *memorized* more number-facts than the rest of us). Even someone (like myself), who is just about guaranteed to make an elementary error of some embarrassing sort when attempting a mental calculation, or even one on paper, isn't regarded as having less of a grasp on the *concepts* of counting, addition, multiplication, and division than those calculational *wunderkinds*. What *does* seem to be required, however, to consider someone as possessing these concepts in a way that's not “defective,” is the ability to recognize a mistake that's made. When an individual's calculations are laid out in detail, and the mistakes pointed out, she should *see* that she's made mistakes, and understand *why* they're mistakes. If not, she's taken as not having fully grasped the concepts in question—or at all.<sup>12</sup>

## 2.2 Why Introspective Resources Are Insufficient for a Response to Meaning Scepticism

I'll start with an assumption I've attributed to Kripke's Wittgenstein. This is that the possible resources of a subject for understanding counting can come—in principle—from only two places. First, from the subject's introspective grasping: her awareness of something that amounts to understanding counting. Second, from patterns of behavior—consciously or unconsciously engaged in—that she's disposed to. It's only after the introspective option is recognized to fail that one turns to a dispositional approach.

Recall the meaning sceptic's challenge. What can be pointed to as indicating that the subject always meant to count and not quount? It's quite natural to first respond with introspective facts. The subject, for example, is straightforwardly characterized as *having recognized* the pattern demanded of counting from having seen a

---

<sup>12</sup>As I'll indicate later, although this recognition-requirement seems to apply to counting and to other simple computational concepts, addition, subtraction, etc., it isn't a requirement on every concept that we take ourselves to understand. On the contrary: It's often thought that the meaning of a concept is understood by someone (or by all of us) even though he (or all of us) may not be able to recognize when some (or even all) of the uses made of the concept are incorrect. (I won't think someone doesn't understand the concept of God even if I disagree with that someone on *everything* or nearly everything he claims about God.)



finite number of cases, and now *seeing* how to go on. But (Kripke (1982, 18)), no finite number of cases determines a unique pattern—a unique way to go on. So too, any attempt to describe the subject as *aware* of a rule or of an algorithm that's compatible with the previous cases upon which she has exercised her counting ability and that determines the answers to future counting tasks she may undertake, fails because any such rule is open to multiple interpretations that can't be made determinate. Any such rule can be reinterpreted compatibly with what the subject "has in mind," and with all her previous counting experiences.<sup>13</sup>

We can now see how Kripke's infinitude requirement rules out a natural description of the needed pattern of grounding facts in terms of what the subject *introspects*. As many philosophers have noted, what causes the problem isn't that the subject has to have acquired an ability to provide right answers for *infinitely many* new cases; it's that the subject has to have acquired an ability to provide the right answers for *new* cases—ones on which she hasn't exercised her counting capacity in the past. Even if a subject faces what strikes the viewer as exactly the same counting task a second time, it's possible for her to do something different and yet describe it as the "same" thing she did before. (The second task, after all, is occurring on a Tuesday instead of on a Wednesday, or during a full moon instead of during a new moon, or it's just occurring *later in time*.)<sup>14</sup>

The problem is that possible occurrent mental states encompassing the understanding of a rule via the couching of those rules in linguistic forms (even ones involving quantifiers), or via couching rules in some other sort of mental imagery (visual, kinesthetic, etc.), as memories of previously executed tasks, etc., must nevertheless be *applied* to a new current problem. But how such rules are *applied* indicates how *they are being interpreted*—and this isn't something determinately fixed ahead of time by what's presently introspectable. Many think this is the lesson of Wittgenstein's (1958) § 139, and nearby sections. Putnam (1981, 20), italics his, puts the point nicely: "What the phenomenologists fail to see is that what they are describing is the inner *expression* of thought, but that the *understanding* of that expression—one's understanding of one's own thoughts—is not an *occurrence* but an *ability*."<sup>15</sup>

Something needs stressing here because of certain responses that have been made to Kripke's discussion. As widely noted (e.g., Wright (1989, 109)), to evaluate

---

<sup>13</sup> Kripke (1982, 7–22). Kripke poses the meaning sceptic's challenge in first-person terms, as whether *my* previous practice and *my* occurrent thinking is compatible with "quaddition" instead of (as presumed) addition. I've instead described the challenge "third-person," and—as noted—in terms of counting (instead of "quounting"). Given that we routinely allow ourselves to describe the phenomenology of third parties—pretty much as I've done above—this raises no complications. The reader uncomfortable with my use of our ordinary practice of describing the phenomenology of others can easily restructure my discussion in first-person terms. It won't affect anything essential.

<sup>14</sup> Kripke (1982, 52, footnote 34) implicitly acknowledges the point when he quotes Wittgenstein: "If I know it *in advance*, what use is this knowledge to me later on? I mean: how do I know what to do with this earlier knowledge when the step is actually taken?" (Wittgenstein (1956, I, §3), italics his)

<sup>15</sup> Wright (1984, 771) is more cautious: "Understanding an expression is, intuitively, more like an ability than a disposition."



whether introspective resources are sufficient to determine that what the subject means at a moment corresponds to what she meant before, Kripke idealizes her conscious access to her past mental states so that she knows *everything* about her former mental life and behavior. Included among her introspective resources, therefore, are her recollections of what she *meant*. It's possible to think therefore that, *of course*, she meant to "count" and not to quount. Perhaps she thought explicitly of the word "count," or perhaps, wordlessly, the concept "count" was involved in her thinking; perhaps there were images, or a grasp of the primitive appropriateness of what she was doing.<sup>16</sup> Why aren't memories like this sufficient to respond to the meaning sceptic? Why can't the subject just respond, therefore, with: "this is what I *mean now* because this is what I *meant before*"?

Kripke (1982, 51) does glancingly reject a version of this option, the meaning-addition-by-"plus"-state being primitive or *sui generis*, complaining that it leaves the nature of the *sui generis* state "completely mysterious." Because of this response (and related reasons), some philosophers have charged Kripke with a bias against anti-reductionist responses to the meaning sceptic.<sup>17</sup> Ginsborg (2011, 229) writes that critics, in support of anti-reductionism about meaning, have objected to Kripke supposing "that the fact of someone's meaning something, or the fact of her following a rule, should be specifiable in purely naturalistic terms."<sup>18</sup>

This underdescribes the scope and methods of Kripke's meaning sceptic. To suggest that a subject simply *grasps* her meanings introspectively, and this suffices to refute the meaning sceptic, really does make these supposed introspective resources completely mysterious. How, after all, do the contents of a subject's mind (public-language words, language-of-thought words, mental states, images, whatever) manage to be about or represent a *specific thing* and not some other thing? One answer some philosophers like is that some mental states (or words, or whatever) just *manage that*, and that's the end of it. In particular there are *intrinsic* meaning states we're sometimes in; and if we're in one of those, and it's a counting state (as opposed to a quounting state), then not only are we counting, but we *know it*.

Alas for this line of thought, what introspection clearly doesn't have access to is a mechanism of *disambiguation*. The intuitive force of Kripke's description of meaning scepticism itself shows that *all* the memories of a subject's execution of certain tasks—memories that *include* the use of words or concepts or images, impressions of what was *meant*, as well as impressions that what she was doing was correct or appropriate—seem compatible with either counting or quounting. The

<sup>16</sup> E.g., Ginsborg (2011).

<sup>17</sup> Boghossian (1989, 527), Fodor (1990, 135–136), McDowell (1984), and others; see Ginsborg (2011, 229), footnote 5 for further examples.

I should add that "rule-following problem" is a misleading misnomer. What's being challenged is *whatever* in the subject is a resource to enable the subject to do subsequently what she meant formerly. The resource in question doesn't have to be a "rule" or a piece of public language with a "meaning." I mention this because some commentators seem to interpret Kripke as working within one or another restriction of this sort.

<sup>18</sup> Indeed, Goldfarb (1985), e.g. 476, reads Kripke as simply engaged in a challenge of "any putative physicalistic reduction of meaning."

problem for a subject, that is, is that her introspective grasping of counting recognizably doesn't have "sufficient content to exclude all the unwanted interpretations of [her] former understanding" (Wright (1984, 765)).

So the charge that Kripke's version of meaning scepticism involves a bias against anti-reductionism misfires. It's rather that the option of a primitive (i.e., non-reducible) meaning-state as a response to the meaning sceptic arises twice, only to fail both times. It fails, first (and now), when investigating the introspective sources available to a subject, and it fails a second time when dispositional approaches are considered. I'll discuss this second failure in Sect. 2.3.

Two last observations before turning to dispositions. First, some philosophers reach for our standard ways of inducing generality from a finite number of cases, and apply them to a subject's access to her own mental contents. On the basis of a finite number of black-raven sightings, we conclude all ravens are black. Why can't a similar grasp of generality arise for a subject when contemplating the memory of her previous executions of counting tasks?<sup>19</sup> Kripke brings against this suggestion an important datum about our experience of meaning to count—and for that matter, meaning anything. *It's not a hypothesis*. This point will arise again as an objection to dispositional approaches; but it also applies here. We simply *know* what we mean; we don't hypothesize what we mean on the basis of an inductive, or any, inference. The second observation is about a related objection to the meaning sceptic that Wright (1984, 776–777) floats:

[Features of our intuitive notion of intention are] available to confront Kripke's skeptic, and ... the skeptical argument is powerless against [them]. The ordinary notion of intention has it that it is a characteristic of mind—alongside thought, mood, desire, and sensation—that a subject has, in general, authoritative and noninferential access to the content of his own intentions, and that this content may be open-ended and general, may relate to all situations of a certain kind.

It can't be doubted this is a correct characterization of what should be called "the phenomenology of meaning."<sup>20</sup> But the meaning sceptic *does* successfully challenge the validity of this experience by asking *what* in this introspective experience provides the needed scope and generality of what's meant—*what* in this experience determines that what's meant is counting and not quounting?<sup>21</sup> Why is it a response to the meaning sceptic that we *feel*—noninferentially—generality and scope in what we mean? Why isn't that we feel the force of meaning scepticism itself proof that we can be brought to recognize that our experience of what we mean is, at best, an

<sup>19</sup> See, e.g., Forbes (1983–1984, 234–235).

<sup>20</sup> Also see Goldfarb (1985) and Ginsborg (2011).

<sup>21</sup> Goldfarb (1985, 474) suggests the Fregean "immediate access to the realm of sense," is a cogent response to Kripke's meaning sceptic. Tait (1986) may be offering the same rejoinder (I'm not sure). In any case, Russell's remark about the virtues of theft over honest toil seems pertinent. We're supposed to be explaining what it is *about the subject* that enables the subject to grasp counting and not quounting. It's no answer to say that certain mental faculties in the subject enable this (and that's the end of it); it's no more of an answer to say that certain objects (functions) are just the sorts of things that subjects *can* grasp (and *that's* the end of it).

introspective *promissory note* the success of which turns on the details of the function intended (and applied)?

## 2.3 How Kripke's Three Constraints Block Dispositional Approaches

It's at this point that one or another dispositional approach is reached for. (Part of the motivation for doing so—notice—is to find resources *in* the subject (*regardless* of whether the subject is aware of them or not) that indicate how the subject is *forced* to the intended interpretation of counting.) I start with a simple characterization of such an account:

DIS 1 To grasp 'counting' is to be disposed, when asked for a number of a collection of objects, to give the number of that collection.

This doesn't work. DIS 1 is compatible with the ability to give the right answers *without* counting. For example, we (adult humans, infants, and many animals) don't *count* groups of one, two or three items: we *recognize* them.<sup>22</sup> It's certainly imaginable that someone might simply *recognize* indeterminately large groups of items as having certain cardinalities (perhaps by the ways that a particular number of objects of various shapes and sizes must fill out space). This person could even have such a skill without realizing—or even being able to realize—that cardinalities of collections can be linearly ordered by magnitude. We must therefore modify the dispositional account to fix on a *particular* way that a subject counts. Here's a candidate:

DIS 2 When someone grasps "counting," and is asked for the cardinality of a collection of objects, he's disposed to point at each item in the collection one-by-one and utter a numeral, starting with "1" when pointing at the first item, followed by the numeral that is the successor of the previous numeral he has uttered each time he points to a new item in the collection, and to give the last numeral he utters when he points at the last item in the collection as the cardinal number of that collection.<sup>23</sup>

Let's see how DIS 2 fares against Kripke's requirements. First, the infinitude requirement. The problem is that people are very likely *not* to be disposed as DIS 2 requires to determine the number of a collection of objects under all or even most circumstances. If someone is tired, or if the collection is too big, the person will fail to do as DIS 2 requires, or refuse to try. This looks like an actual fact about all people: the dispositions DIS 2 purports to describe are real ones. As such, their dis-

<sup>22</sup> This is "subitizing," and it seems to be restricted to about four items for adults and to three items for children, and for certain animals. See Mandler and Shebo (1982).

<sup>23</sup> DIS 2 illegitimately singles out a *single* method. As noted in Sect. 2.1, a subject may have acquired other appropriate short-cut methods of counting. I set aside attempts to refine DIS 2 to handle this because such refinements won't affect the overall dialectical trajectory of this chapter.

positions are sensitive not only to the cardinal sizes of the collections that the subject is faced with, but also to the properties of the items in those collections as well as the background states of the subject. For example, if some of the items reflect sunlight too brightly, or are too close together, or are quite tiny, or crawl around on each other, the subject may actually be disposed to drop the task altogether (and run for her life) rather than engage in soberly correct counting. The upshot is that, according to DIS 2, *no one* grasps counting (because *no one* is disposed, perhaps in the majority of counting tasks, to do as DIS 2 describes).

As Kripke (1982, 30–32) notes, one can instead characterize the disposition in terms of what someone *would do* if their brains were larger, or their eyes not so sensitive to reflected light, or they were less fearful ... One may easily find, however, that such a counterfactually-modified DIS 2 nevertheless gives wrong answers. It may be that whatever changes *in persons* are needed so that they (counterfactually) succeed in counting collections of objects that they're currently disinclined to undertake or manage correctly will instead cause them to deviate from executing the task correctly. There is no way we can guarantee that the changes in persons, to ameliorate the cause of their tendency to avoid or incorrectly execute counting tasks (in particular cases), don't affect them in other ways that continue to prevent execution of those tasks.<sup>24</sup>

Now let's bring in Kripke's third constraint. It might be thought that what's (really) wrong with DIS 2 is that it doesn't accommodate, as Kripke's third constraint requires, the empirical fact that we make mistakes. We don't, that is, ordinarily demand of those we take to understand counting (or ourselves, for that matter), that they on any occasion be able to count successfully. We allow exceptions for various reasons including, but not restricted to: the subject becoming tired during the task because of the large size of the collection, the items in the collection having properties that distract the subject from the task, or the subject being ill or drunk. We even allow *inexplicable* failures by the subject to complete the task correctly. ("Oh wow—*what* was I thinking?") Despite deviations in these and other cases from the behavior required by DIS 2, and even in a subject's dispositions to behavior, we're still willing to describe him as understanding how to count. Why not, then, simply exclude such cases in "exception-clauses," so that a subject is regarded as grasping counting despite the deviant behavior?

The suggestion begs the question against alternatives. Consider succession\*, which differs from succession in a number of cases. We can, in exactly the same way, describe the subject as understanding how to count\* by taking the number-counting tasks where the subject tends to deviate from counting\*, and placing *them* in exception clauses. What, then, tells us that people grasp or understand counting instead of it being counting\* that they grasp or understand?

This suggested fix could have been raised earlier to address the failure of DIS 2 to meet Kripke's infinitude requirement: exception clauses could have been introduced

---

<sup>24</sup> The *actual* details of our biology—e.g., brain organization—physical laws, and so on, are pertinent to the truth of the counterfactuals involved. See Fox (2011) for a popularized discussion of the issue, with pictures and diagrams.

to also handle the infinitely many cases in which subjects are disposed not to engage in the assigned task.<sup>25</sup> But it similarly begs the question against alternative approaches, DIS 2\* say, where “qucessor” is substituted for “successor” in the formulation of DIS 2, and the two notions differ on the answers that subjects don't give.

The problem with DIS 2 is that it attempts to identify the dispositions of a subject (when faced with a counting task) with the *correct* procedures for counting. But as we've just seen, it's no better to characterize the dispositions in question as those that actual subjects have, but supplement that characterization with exception clauses where those clauses deviate from the correct answers. Defending—in this way—the suggestion that the subject understands counting, fails. Proponents of the claim that the subject understands some nonstandard form of counting can use the same tool.

Let's turn, lastly, to Kripke's justification requirement. The problem, simply, is that the dispositional view offers resources that aren't the ones that subjects use (or can use) to justify that in the past they understood or meant themselves to be counting. Kripke (1982, 23) writes:

Am I supposed to justify my present belief that I meant [counting], not [quounting] ... in terms of a hypothesis about my past dispositions?

I read his point as this: the *mere* description of one's dispositions to explain that one *meant to count* fits poorly with the phenomenological impression we have of how people understand their own grasp of concepts such as counting. Listen to how odd the following speech would be:

I know how to count because I'm always disposed to answer the question “what is the number of items in that collection?” by pointing at each item in the collection one-by-one and uttering a numeral, starting with “1” when I point at the first item, and uttering the numeral that is the successor of the previous numeral I uttered each time I point to a new item, and to give the last numeral uttered when I point at the last item in the collection as the number of that collection.

If someone said this, she would raise the suspicion that she *doesn't* understand counting, but has instead memorized a procedure that she doesn't grasp at all. Or perhaps that she's describing an *urge* she feels rather than a course of action she takes under the guidance of a concept that she understands.<sup>26</sup> And we would suspect this because her little speech doesn't indicate anything about why she thinks a description of her dispositions is even *relevant* to whether she understands how to count.

I mentioned earlier that we often prove to someone that we know how to do something by exhibiting the appropriate behavior—for example, by counting a collection: “You think I don't know how to count? Well, watch this!” But *this* person

---

<sup>25</sup> The counterfactual suggestion and the exception-clause suggestion amount to the same strategy executed with grammatical variations. In the former case, diagnoses of the violations of DIS 2 are built into antecedents of counterfactuals; in the latter, a modification of DIS 2 contains the exceptions in antecedents of indicative conditionals.

<sup>26</sup> Kripke (1982, 17) writes: “Normally, when we consider a mathematical rule such as addition, we think of ourselves as guided in our application of it to each new instance.” Also (Kripke (1982, 10)), “I follow directions ....”

isn't offering her disposition to do such-and-such as a *criterion* of what counting is. Rather, her providing such illustrations to another person indicates that she presumes to already know that the other person knows what counting is, and she's proving to such a knowledgeable person that she knows *this*, and also knows what counting is. Both are achieved by the illustration of the pertinent skills she has (and by her indication that she has them) *to* the other (knowledgeable) person.

Notice that the stress I interpret Kripke as placing on how the justification requirement undercuts dispositional approaches isn't that justification is "normative" and dispositions aren't. It's, rather, that the way in which subjects understand themselves as justified isn't satisfied by dispositional approaches. Dispositional approaches violate facts about how subjects *experience* themselves as justified.

## 2.4 An Attempted Defense of Dispositional Approaches to Meaning Scepticism

Many of Kripke's commentators defend the dispositional approach by exploiting an apparent analogy between the use of counterfactuals to describe the behavior of glass, salt, or gases, in nonactual circumstances, and the corresponding use of counterfactuals to describe what a subject would do under certain idealized circumstances—e.g., having a larger brain, being immortal, well-rested, etc.<sup>27</sup> They suggest that such counterfactuals—about glass, salt, and so on—involve both the exception clauses (*ceteris paribus* clauses) and the idealizations that fail, according to Kripke, as a response to meaning scepticism. Furthermore, glass, salt and gases face analogous versions of Kripke's infinitude requirement, and those versions are handled dispositionally. In infinitely many different circumstances, the structural facts underlying the dispositions of glass and salt determine open-endedly many exception clauses—salt *won't* dissolve in cool salt-saturated water, salt *won't* dissolve in water if God intervenes to adjust the motion of each molecule in the particular way required by *that* specifically-shaped molecule in order to prevent ....

There is an important disanalogy, however, between glass, salt, and the like, and a counting subject.<sup>28</sup> This is that the actual structural facts that underwrite the idealized descriptions of glass, salt, and gases, and the like, are taken to determine the ways in which these idealized laws are *false*, not true. It isn't, after all, that salt *does* dissolve in water under all circumstances, or that gases behave just like collections of little rigid round spheres. More importantly, what salt is *disposed* to do (under

<sup>27</sup> Blackburn (1984, 289–290), Forbes (1983–1984, 229–230), Fodor (1990, 94–95). Others applaud this response, e.g., Boghossian (1989), Ginsborg (2011). Wright (1984, 771) rejects the suggestion tersely on the grounds that, intuitively, understanding is more like an ability than a disposition.

<sup>28</sup> What follows in this and the next five paragraphs, although a defense of Kripke's rejection of dispositional approaches, isn't (as far as I can tell) given by Kripke. (Or, for that matter, by anyone else.)

various circumstances) is exactly what salt *does do*—based on its chemical/structural properties. Salt is *not* “water-soluble”—understood *literally*, this is false. Salt, rather, dissolves in water in a broad range of *such-and-such circumstances* (but it doesn’t dissolve in water in *so-and-so* other circumstances). The circumstances in which salt doesn’t dissolve in water, however, are special enough that the label “water-soluble” is nevertheless useful. Similarly, it’s that gases—under a wide range of circumstances—*approximately* behave the way media composed of little rigid round spheres would behave (if there were any).

Counting is different. In studying the neurophysiological basis of numerical counting abilities, *it’s allowed that* we may discover that subjects are *disposed* to do something other than count although *it’s not allowed that* we’ll discover that subjects actually don’t count (don’t grasp counting) but instead do something else. It’s empirically allowed that we might discover that a quounting function correctly characterizes human dispositions; this would *not* be taken, however, to be a discovery that subjects quount, and don’t count. It would not be taken as showing that subjects don’t grasp counting—don’t know what counting is—but instead understand and undertake quounting; it would, instead, be regarded as a discovery that subjects—all subjects—are disposed to make certain systematic errors when they count.

I’m *not* making a prediction based on a thought experiment; I’m describing current research in the study of mathematical abilities, and what discoveries about human (and animal) dispositions are understood in that research to *reveal*. The first point to make is that, *in fact*, our discoveries about mathematical dispositions in subjects show that they deviate from what we take correctly executing those mathematical tasks to demand.

I’ll briefly expand on this. Philosophers have suggested that the dispositions needed to respond to Kripke’s sceptic can be “stratified” to separate primary dispositions to correctly count from secondary dispositions that “interfere” with the primary ones.<sup>29</sup> Although this distinction corresponds somewhat to our experience of mathematical tasks (we make mistakes that—*sometimes*—we recognize as due to interfering factors), this isn’t a property of the dispositions involved in our numerical competences. Instead, what’s been found is that our capacities to engage in numerical tasks are constituted by several subpersonal cognitive systems that are intrinsically limited in the tasks they can successfully perform.<sup>30</sup> One interesting fact is that our intuitive number line obeys Weber’s law: the numbers are spaced

---

<sup>29</sup> See, e.g., Blackburn (1984, 290) and Forbes (1983–1984). Goldfarb (1985, 477) writes: “A reductionist could claim, for example, that future physiological psychology might reveal two mechanisms, separable on scientific grounds. States of the first amount to a person’s linguistic competence, and would, if untrammelled, always cause correct responses; states of the second are identifiable with interfering features, which explain why on particular occasions the first mechanism does not issue an appropriate response (and the person errs).” Although, *in principle*, infinitary dispositions that generate the *correct* functions are neurophysiologically possible—this was ruled out by the early 1990s as far as arithmetical competences are concerned.

<sup>30</sup> Some relevant literature: Carey (2009), Dehaene (1997). There is much more, of course. This is a very active research area, primarily because the “dispositions” being studied are so intricate and



unevenly, bunching up as their magnitudes increase. We also have a fuzzy analogue sense of the numbers of groups of objects.

Were our dispositions constitutive of our understanding of numbers, arguably, we would have more than one system of numbers that we understand, when we undertake counting tasks; one is fuzzy, but open-ended, another is strictly finite, a third is one in which distances between magnitudes obey Weber's law. It's certainly a requirement of any proponent of a dispositional approach to our grasp of counting to explain how this bewilderingly complex welter of neurophysiological capacities underwrites a simple stratified grasp of standard arithmetical functions coupled with *ceteris paribus* clauses to disallow as relevant to our number concept behavioral deviations from those arithmetical functions.

This leads to a second devastating point about empirically-discoverable dispositions. Even if the dispositional complexities that I just described weren't being discovered by current research, the fact that we recognize it's an empirical possibility that subpersonal arithmetical capacities deviate from what we understand as the correct execution of arithmetical tasks shows that we don't take the understanding of counting (and our understanding of other mathematical tasks) to be *constituted* by our dispositions to execute those tasks. It's in *this sense* that a distinction between what we are inclined to do, and what we *ought* to do—what's *correct* to do—should be drawn.

This does raise an issue. Given that dispositions—the items studied by the appropriate scientific researchers—exhaust a subject's resources for understanding counting, and other mathematical concepts, what then determines that subjects mean to *count*? This may be described as the question: what fixes the meaning—the content—of a subject's concept of counting? What *could* fix it, after all, if not a subject's dispositions?<sup>31</sup>

There are two observations to make. The first is that the issue of content fixation is coming up (here) in a dialectical search for *something* in the subject that determines that the subject grasped counting in the past, and not, instead, quounting; and consequently, for something that determines that the subject currently grasps counting and not quounting. We could have run a challenge to what a subject means directly off of the failure of dispositional approaches (leaving aside the considerations of Sect. 2.2)—but that would require additional assumptions I'm not interpreting Kripke as presupposing. Kripke's meaning sceptic accompanies arguments for the failure of dispositional approaches with arguments against introspective resources.

Second, this is a way that either arithmetical laws or our grasping of such laws could be described as “normative.” After all, what the failure of dispositional approaches amounts to is that the fact that “one is or was disposed to respond in a

---

gerrymandered—not a simple matter of embodied arithmetical rules that people mechanically execute unless something interferes so that they make blunders.

<sup>31</sup> Content fixation is the issue Boghossian (1989) takes Kripke to be directly raising; it's the issue Wright (1980) does directly raise (see footnote 4). But the considerations Boghossian uses to motivate this issue aren't the ones I've just given, for he (2003, 496) writes, describing his earlier views: “I used to underestimate the force of [the infinitude requirement].”



certain way on a given occasion cannot make it the case that one ought so to respond” (Ginsborg (2011, 228)). It’s better—less misleading I think—to say directly what (and all) this means: only the negative point that a description of dispositions in a subject doesn’t characterize that subject’s grasp of counting. The words “correct” and “ought” are *solely* driving a wedge between a description of the possible behavior of a subject, and the actual function (e.g., the successor function) taken as grasped by that subject. A certain “reduction” fails; that’s *all*. Similarly, to say, for example, that “correctness is a normative matter” (Boghossian (2003, 36)) here means nothing more than the actual successor function isn’t what the subject is *disposed* to execute.

It’s tempting, I think, to suggest that Chomsky’s old competence/performance distinction stands or falls with the primary/secondary distinction in dispositions that I just undercut. Not so. One reason is that both sides of the competence/performance distinction are empirically sensitive: “performance” isn’t a mere trash bin for whatever ungrammatical behavior the subject exhibits. The syntax of the language attributed to the subject is revisable because of empirical results; also, the suggestion that the subject is suffering from a performance failure is empirically constrained, and empirically testable. Memory failure can’t simply be deduced as a reason for why the subject fails to exhibit the correct syntax in speech; that memory failure is involved is itself open to empirical testing. A second related reason for the disanalogy is that a major factor in the arithmetic case is missing in the syntactic case: correctness. As I indicated earlier, universal dispositions to execute counting tasks according to deviant (finite, fuzzy, etc.) number systems aren’t indications that subjects don’t count; they *would be* evidence that the violated syntax isn’t as the linguist hypothesizes it.

## 2.5 The Failure of a Straight Sociological Solution to the Rule-Following Paradox

As we’ve seen, individual dispositions are severely limited in their ranges; they don’t meet Kripke’s constraints. In other words, they can’t provide the needed psychological/individual grounding facts for every (possible) counting task, where grounding facts are understood as sufficient for determining correspondence facts such as: *Every collection of objects has one and only one cardinal number*. This is what makes it tempting to replace someone’s actual dispositions with counterfactually-supplemented dispositions. Doing so is a very intuitively-natural response to the problem of limited powers because we often excuse ourselves by saying that if we hadn’t gotten tired, we wouldn’t have made certain mistakes, or that we would have (successfully) undertaken a counting task that we have otherwise had to decline. In this way, we abstract the characterization of our knowledge of counting, and our ability to undertake it, from other factors that limit our execution of counting tasks in much the same way that the actions of Turing machines are

abstracted away from limitations of physical machines, tape, and energy. This is what gives intuitive force to the idea that Kripke's infinitude constraint can be met by a counterfactual supplementation of our actual dispositions.

But this strategy won't satisfy Kripke's justification constraint. And, indeed, that it's even cogent to worry about whether the supplemented dispositions can do the job needed shows that the dispositions—*whatever they are and however they're supplemented*—aren't the standards by which correct counting *should be* judged. If they were, there could be no question: *whatever* answer dispositions yielded would be understood as correct. But, regardless of what one's dispositions to count look like, it seems that anyone can sensibly pose the question of whether his dispositions give the right answer; more strongly, it seems that it's sensible *to* pose the question of whether one's dispositions give the right answer. It's bizarre for someone to say, when wondering if he has given the right answer after a count: "Well, it's the right answer because it's the answer I was disposed to give."<sup>32</sup>

This invites looking elsewhere for the standards by which correct counting is to be judged. Contemporary scientific practice, furthermore, suggests that a focus on the dispositions of individuals to count is, in any case, largely irrelevant because the ability of *individuals* to count is one that's very nearly never relied on. Most of the counting results we have aren't based on any individual's capacities, but instead on the use of scientific theories and instruments. Call the instruments involved "calculators." Given the widespread use of calculators, it almost seems like the only relevant disposition that (most) individuals need is an inclination to accept results due to others, or due to calculators; all that's required, that is, seems to be a disposition to *defer*. Indeed, the historical use of counting tools such as abacuses, counting boards, or knots on strings, indicates that something like this has been the case for many centuries. Our capacity to determine the numbers of many collections—and to estimate these numbers—doesn't turn on anything reasonably described as the dispositions of individuals to count (unless, as just noted, we regard it as part of someone's disposition to count that he'll accept the results of instruments—calculators—that other people have invented and developed).

I stressed in Sect. 2.1, while discussing Kripke's justification requirement, that individuals feel justified in their choice of methods that they use to count. This is even true of the (confident) child who will simply be puzzled by the adult's claim that everything she does is compatible with her knowing how to quount. Kripke uses this phenomenological fact against the dispositional account with great success because any subject's belief in his being so justified to do what he does isn't satisfied by the mere recognition on his part that he's disposed to undertake a process with such-and-such characteristics. This phenomenological fact, coupled with

---

<sup>32</sup> This is an extremely important aspect of how we understand rule following, and it's due to relatively deep facts about our (shared) psychology—in particular to the peculiar ways that we take account of (and fail to take account of) the subpersonal abilities that are relevant to consciously undertaken tasks. This interesting interplay between the personal and the subpersonal has no role in my initial models of private rule following (my first two *Robinson Crusoes* in Chap. 3); but is part of the realistic psychological aspects of Crusoe 5, to be described in Chap. 5.

additional phenomenological facts, may seem to imply that it's social standards against which correct counting behavior is to be judged.

Consider, first, that the subject's confidence in his knowledge of counting is frail. Even if an individual is very confident of his abilities, the right external circumstances can make him quite unsure of what he thinks he knows.<sup>33</sup> This is not only because it's a fundamental part of one's understanding of any such task that one is "capable" of mistakes, and that these mistakes are seen as correctable by others or by instruments. It's also because a fundamental part of one's understanding of any such task is that the magnitudes of the mistakes (that one is capable of) have no lower limit. One can be brought to become unsure of *anything* one thinks one knows. If how one was disposed to act or respond were *itself* the standard of one's behavior, this insecurity would make no sense. And *this* apparently shows that the standards in question are external to any single individual, and located in the people around that individual—in her community.

Indeed, it can be suggested that the "normative" elements that Kripke mentions as part of the phenomenology, and that bear so negatively on the dispositionalist response, namely that (as cited in footnote 26), "Normally, when we consider a mathematical rule such as addition, we think of ourselves as guided in our application of it to each new instance," and "I follow directions," point pretty unequivocally towards a large part of our impression of being guided in the case of counting being that we feel we have grasped a rule *that others have taught us*. That is, that it's part of the *phenomenology* that the source of our confidence in what we do is that we take ourselves to have learned it correctly (from others). And this explains why experiments can be so easily designed that undercut this confidence. Everyone else isn't agreeing with me; and one possibility is that I haven't learned what I thought I learned—or that I've made a mistake—or worse, that I'm now experiencing the psychological effects of a medical condition.

The location of standards in the (dispositions of the) community one belongs to will also explain the result of a thought experiment, say, where someone becomes conscious of her subpersonal source of her counting ability. Suppose she suddenly starts to see bright red visualizations of numerals that are associated one-by-one with each object in a collection she's counting, and where the last one to so be associated with an item is the number she feels an *urge* to utter as the number of the collection.<sup>34</sup> The natural reaction to this change in one's (conscious) psychology is the fear that perhaps the answers so urged *aren't* justified. This fear arises because of the apparent detachment of the urge-mechanism from the way we think of our-

---

<sup>33</sup> A routinely-described psychological result is the insecure reaction of a subject in a situation—e.g., in a classroom—where very simple instructions have been given ("raise your hand when the teacher holds up a green card, and lower it when she holds up a red card"), when, after a few episodes of this, everyone else in the class (as previously arranged) conspiratorially violates the instructions.

<sup>34</sup> What kind of urge? Imagine an urge like the ones we have (for example) to scratch an itch, or urinate, or eat. It's resistible, but with (varying) difficulty.

selves as having learned how to count. We learned it from others; but how can we be sure that the urge we've now discovered to be operative in us conforms to what we're supposed to have learned?

These considerations invite a straight *sociological* solution to the rule-following paradox. There *is* a pattern of grounding facts that indicates that the child is counting, and not quounting, but it's found in the collective dispositions of the entire society within which that child has learned to count. Kripke (1982, 111) raises this option in passing, and notes that such a theory "would be open to at least some of the same criticisms as the original [individual-disposition theory.]"

Indeed, the straight sociological solution fails to satisfy Kripke's three constraints in *exactly* the same ways that the individual-disposition theory fails to do so. *The infinitude constraint*: The collective disposition of an entire community—however that's defined in terms of collections of individual dispositions and collectively-available instruments—is still limited in its range, and still capable of performance deviations from correctly-applied counting events. This is true even of *our* contemporary community—with its powerful computing instruments. Counterfactual supplementation of the actual dispositions of the community faces the same obstacles it faced in the individual case. *The justification constraint*: A description of the dispositions of the community at a time doesn't indicate why that community is justified in computing new cases of counting the way it does—that by virtue of its (collective) dispositions the counting-concept it exhibits in the new case should be taken to be the same one that it previously exhibited in earlier cases. We do *not* think—even when we conceptualize counting tasks as ones we undertake collectively—that the way we do it is (by definition) the way it should have been done, given the way the community-practices were understood to operate previously. Finally, *the mistake constraint*: a collective solution offers no logical space for the required possibility of there being mistakes in calculation, not just by individuals, but even by an entire community. Attempts to meet Kripke's mistake requirement for collective dispositions are open to the same question-begging concerns that were discussed in Sect. 2.3.

## 2.6 Assertability Conditions, Not Truth Conditions

Let's turn, therefore, to Kripke's Wittgensteinian sceptical solution. A sceptical solution (Kripke (1982, 66)) concedes "that the sceptic's negative assertions are unanswerable," that is, it concedes that there is no pattern of grounding facts that can be used to answer the question of whether a subject is counting or quounting. Wittgenstein's solution (so Kripke argues) is to desert "truth conditions" for sentences about a subject's understanding of numeration statements, and indeed "truth conditions" altogether. Because there is no pattern of grounding facts for someone's understanding of true sentences (about numeration) that can meet Kripke's three requirements, there is nothing that determines necessary and sufficient truth conditions for sentences about a subject's understanding those (numeration) statements.

Instead, one substitutes “assertability conditions.” “Assertability conditions” are the giving of conditions for when one is *licensed* or *entitled*—in a community—to assert (or deny) of oneself, or of others, that they understand counting. It’s built into the notion of assertability conditions that a person belongs to a community in which such conditions are (recognized to be) in place.<sup>35</sup>

Kripke suggests that the assertability conditions for A’s entitlement to the claim that he means addition by ‘plus’ is that (roughly) he’s confident that he can give correct answers to new cases, subject to correction by others. And that he’s also entitled (Kripke 1982, 90), “again provisionally and subject to correction by others,” to judge a new response by someone else as correct if that response is the same response that he’s inclined to give. A denial that another understands counting is entitled if that person’s answers deviate from one’s own, and a denial of one’s own understanding is entitled if one experiences insecurity about what the answers are or about how one is to go about finding them. Kripke (1982, 96) stresses that such assertability conditions have a point in our society because (most) individuals in our society have similar dispositions.

As indicated above, “the community” is relevant in two ways to these suggested assertability conditions. First, a reference to the community occurs in the statement of the assertability conditions themselves: We are taken, that is, to recognize the provisional nature of our dispositions—that it’s legitimate for *others* to correct us. Second, the community is relevant because what’s on offer are assertability conditions; and assertability conditions themselves only make sense in the context of a *community* because they rely crucially on the notion of entitlement or license in a

---

<sup>35</sup>I take myself (in this section) to be expounding Kripke’s (1982, 74–93) discussion of assertability conditions. Douglas Patterson (9/24/09—email) has raised the concern that a general replacement of truth conditions (for contents or statements) by assertability conditions leads to the view that it’s *sentences* “otherwise meaningless, that have assertability conditions.” And so, “what I’m entitled in the community to do is to make a noise, say.” He further suggests that such a view isn’t coherent unless an illicit appeal to truth conditions occurs in the description of what people are licensed to do.

I think this is wrong—despite my eventual denial in this book that truth conditions need be replaced by assertability conditions to provide a sceptical solution to the rule-following paradox. I agree that sentences (without necessary and sufficient conditions of application) are what—on this view—people are entitled to apply or refuse to apply. Two views are possible: (i) The pattern of entitlements of such sentences *is* their “meaning”; (ii) Because of certain conditions that “meanings” must meet (e.g., that they must result in necessary and sufficient conditions of application) and that such a pattern of entitlements doesn’t meet, these sentences don’t have “meanings.” In neither case, it seems to me, does it follow that an assertion practice (built on “entitlements”) is incoherent, or that it needs supplementation with truth conditions.

The replacement of truth conditions by assertability conditions—Kripke (1982, 86) describes Wittgenstein as claiming—doesn’t rule out ordinary uses of “true” and “false.” For exactly these reasons, as I indicate in Chap. 6, one can still utilize—instead of assertability conditions—truth conditions, provided one understands them in a “deflated” way, rather than as requiring correspondence relations to facts. I’m suggesting, in other words, that Kripke (and perhaps Wittgenstein) has built into the notion of “truth conditions” that they are required to provide correspondence relations to facts (Kripke (1982, 72): “A declarative sentence gets its meaning by virtue of its *truth conditions*, by virtue of its correspondence to facts that must obtain if it is true” (italics are Kripke’s)). I deny that “truth conditions” need to be so understood. See Chap. 6.

community to assert something or to deny it. Such concerns with entitlement or license lapse as relevant for isolated persons outside a community—what I'll call *Robinson Crusoes*.

According to Kripke's (1982, 86–88) interpretation of Wittgenstein, the incoherence of a lonely rule-following Robinson Crusoe is a corollary of the fact that such an isolated figure can't function in a context with only assertability conditions. When he takes himself to be counting, he acts "unhesitatingly but *blindly*" (Kripke (1982, 87), *italics his*). The sceptical paradox tells us that if we restrict our attention to Robinson Crusoe alone, there are (Kripke (1982, 89)) "no truth conditions or facts in virtue of which it can be the case that he accords with his past intentions or not." Thus, there is no difference between Crusoe thinking he's following a rule and his actually following that rule; and therefore the idea collapses that one can correctly or incorrectly follow a rule that one previously intended to follow. As Wittgenstein (1958, § 202), writes:

To think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately'; otherwise thinking one was obeying a rule would be the same thing as obeying it.

Kripke (1982, 110) stresses that it *doesn't* follow that a Robinson Crusoe can't be described as following rules (for counting) correctly or incorrectly; it's that for someone to describe Crusoe this way requires that person to treat Crusoe as belonging to his (the describer's) community.

Notice how the failure to provide a ground for Kripke's second, and especially third, requirements on any solution to the rule-following paradox, when restricted to the dispositions of the individual subject, is brought directly against the suggestion that a Robinson Crusoe can—independently of a community—count. Notice also Kripke's suggestion, that when we judge a Robinson Crusoe is counting, we incorporate him into our community (and apply our assertability conditions, with our accompanying standards), follows directly from this being the only way that the required notions of justification and mistake (provided by Kripke's Wittgensteinian assertability-conditions response to the paradox) are enabled to operate.<sup>36</sup>

*Final Remarks* Kripke's masterful exposition leaves most readers with the strong impression that the logical contours of this problem area *force* Kripke's sceptical community solution to the rule-following paradox. Kripke's three constraints on grounding facts arise from simple observations about our rule-following practices and those constraints pretty plainly rule out (straight) dispositional solutions—as this chapter has shown. In turn, correspondence facts for our assertions are undermined, not by metaphysical results that directly undercut the capacity of our assertions to describe (or misdescribe) what's out there, but instead because *we* lack the resources to make our assertions so correspond. The point, to use the counting-example, isn't anti-Platonistic—that there isn't a successor function. It's that the

---

<sup>36</sup>A distinction is drawn between "solitary languages" spoken by single individuals and Wittgenstein's "private languages," e.g., "sensation languages" (Goldfarb (1985)). Blackburn, Goldfarb and Boghossian think Kripke's skeptical solution doesn't rule out "solitary languages" as impossible. I don't agree.

subject's ways of thinking and speaking *about* that function are undermined by the absence of the needed grounding facts. The absence of correspondence facts then follows: the mental items of the subject or of her language can't correspond to what we (pre-theoretically) take them to correspond to because a subject is unable to mean her sentences or thoughts to so correspond. (This is an extremely important theme that arises again in later chapters—specifically in Chap. 4, which is focused on the Lewisian denial that correspondence facts are based solely on the subject's resources.)

The absence of correspondence (given natural assumptions about the properties of the truth idiom) in turn undermines truth-conditional analyses of our assertional practices, and this requires their replacement by an assertability-conditions analysis. Natural psychological facts, finally, about our incapacity to place standards on ourselves (ones that require us to determine all by ourselves whether we meet or don't meet them) force the intrusion of community standards to make rule following cogent. The dialectic just rehearsed is *very* compelling.

The dialectic is also very broad in scope. It's couched in terms of arithmetical functions—where “following a rule” most naturally seems to apply. But, as Kripke notes, the dialectic is equally compelling for “meaning to” apply a word or use a concept or just meaning to do what one did earlier—all cases where rules (strictly speaking) needn't be operative.

I show in the succeeding chapters of this book how to deflect these conclusions.



## Chapter 3

# Two Versions of Robinson Crusoe

**Abstract** This chapter begins the analysis of the rule-following problem by the use of disposition-languages, languages with terms that apply exactly the way subjects are disposed to apply them. It's shown that if empirical circumstances (and the dispositions of the subjects) are felicitous enough, isolated subjects can engage in successful rule-following. That is, they can successfully evaluate the languages they speak as better and worse, and they can use these languages to successfully navigate their worlds. Two examples of rule-following isolated individuals are given, Crusoe 1 and Crusoe 2. Their dispositions are somewhat artificial. Later examples, in later chapters, of isolated rule-following individuals will be much more natural, that is, they will be much more like us.

### 3.1 A Community of Idiolects

One result of this book is that private models of rule following *are* possible. I show this in stages, using a series of differing Robinson Crusoes all of whom can engage in various kinds of private rule following. To prepare for this series of thought experiments, I first briefly present in this section an imagined community that, contrary to the practices of *our* community, *does* fix the meanings of their words in terms of their dispositions. I'll explore some of the elements that cause such an alternative language practice to break down and I'll illustrate some of the ways in which, perhaps surprisingly, such a practice nevertheless succeeds.

*Disposition-meaning Languages* Imagine a community in which each individual really does intend each word that he uses, "table," "count," "add," "owl," and so on, to mean in each context exactly what he has an urge at that moment to apply that word to. That is, by "dog," a speaker means those items he has an urge to characterize as "dogs," by "table" those items he has an urge to apply the word "table" to, and so on. One and the same object, can, as a result, be a dog in some circumstances but not in others where it might be a "cat." In most circumstances, of course, it isn't anything at all. (The extensions of words in these languages aren't sets of objects but sets of object-context pairs, where contexts are individuated by tense, a portion of location, and other standard factors.)



It may seem that this suggested language trades on an inherent circularity that's fatal. After all, the meaning of "dog" is given by a formulation involving the very word "dog." It's true: there *would be* circularity if the people in this community understood these formulations as providing them with *instructions* for using these words. But that *isn't* their understanding. Their dispositions—even if acquired in learning contexts—are, and are recognized to be, automatic procedures that *dictate* answers; and so the formulation for the meaning of "dog" independently characterizes its meaning by treating it as applying to those items precisely when an urge to so apply "dog" is operative. Call such languages "disposition-meaning languages," and describe their words as having "disposition meanings."

Kripke's three requirements lapse or are trivially met for these languages. The justification requirement is trivially met because the terms of the language used at a moment are sensitive only to the speakers' urges at that same moment. (What speakers mean then and there is what they feel urged to mean then and there.) Speakers have no intentions (although they *may* make predictions) about what they'll mean in the future, and at each moment they don't regard themselves as conditioned by what they meant in the past. So if the sceptic asks, "How do you know that what you meant then is what you mean now?" the response is:

I do remember meaning by "cat" what I felt urged to mean then. And I remember what it was I applied the term to. If you're asking, "do I still mean at this moment what I feel urged to mean at this moment, just as I did at past moments?" the answer is 'yes'. After all, I remember perfectly clearly what happened in the past. And if you're not asking that, I don't know what you're asking.

The sceptical challenge can't get started because the invocation of someone's then dispositions *really does* justify what he meant then, just as he's currently justified that his current behavior is in accord with what he means now. The bizarre speech of Sect. 2.3 isn't bizarre for anyone in *this* community. Another way to put the point about the failure of the meaning-scepticism challenge to speakers of these languages is this: although there *is* a generalization that applies to what speakers of these languages mean, *they always mean what they feel an urge to mean*, this generalization isn't part of what any speaker intends when they mean something. They don't *intend* to always mean what they feel urged to mean, they simply always do.

The mistake requirement lapses because a person's application of a word is always correct no matter what her dispositions (and behavior) are like. One cannot perform "badly." Lastly, no one in this community intends, in any case, to "add" or to "multiply," or to apply any words, for that matter, as *we* understand ourselves to so intend: each only intends to perform "addition" and "multiplication" as her activities will be executed given her own dispositions at each moment. So only if someone's dispositions allow her to mean a notion of "counting" that applies to arbitrarily-large collections will her notion so apply.

One possible drawback of this disposition-meaning "language" is that, unless the people in this community have *exactly* the same dispositions (in exactly the same contexts), they won't have words in common; if there are deviations in their dispositions—however small—with respect to many words, the result will be a family of

individual idiolects rather than anything that can be described as a “public language.” Furthermore, (i) they will recognize this fact, and (ii) they will recognize that there can be no translations from one idiolect to another (at least as far as a large class of words is concerned). Under such circumstances, their cacophony of idiolects might seem to be useless for communication. These will be *seriously-private* languages!

Such a pattern of idiolects won’t be useless for (most) communication, however, if the idiolects deviate from one another in only small ways. For in that case communication failures will only occur sporadically. In general, therefore, how successful members of the population will be at communicating with one another will turn on how *much* deviation there is among their idiolects. Apart from this, and even if *almost* all their words involve substantial interpersonal deviations, communication can succeed under certain circumstances: if translations exist.

Here’s one way that could happen: Suppose B—a member of this community—has the word “disposition”; and suppose that her word applies *exactly* to C’s dispositions and to her own. In such circumstances, B can recognize that how she means to understand her own words, nouns such as “table” and “orange,” verbs such as “run” and “multiply,” and so on, is via her own dispositions to apply such words. B similarly recognizes that this is true of C. B further recognizes that she can’t translate C’s idiolect into her own directly: “table,” as C uses it, doesn’t have the same application conditions as B’s word “table” because their dispositions differ. Nevertheless, B can still translate C’s word “table” like so: “table,” as C uses it, means in B’s own idiolect, “‘table’ as C is disposed to apply ‘table’.” If everyone’s word, “disposition” so applies to others’ dispositions—if (this is a corollary of this assumption) they share the word “disposition” and if that word’s application conditions exactly fit everyone’s dispositions—then mutual translation of everyone’s words in this community is possible.

Of course, we can imagine that *all* the words of all the speakers in this community deviate, including “disposition.” And, further, it can be that the way that B uses “disposition” causes the phrase “‘table’ as C is disposed to apply ‘table’” *not* to have the same application-conditions as C’s “table” (because B’s use of “disposition” doesn’t correspond to C’s dispositions). This, if recognized, enables B to see that C and herself speak different idiolects that aren’t mutually translatable to one another. Whether, of course, despite the impossibility of translation they can still communicate with one another turns—as noted earlier—on how different the application conditions of their respective words are.

As the foregoing indicates, the impossibility of translating these idiolects to one another, and—more generally—the failure of such languages, at least for purposes of communication, doesn’t inexorably follow from a wedding of the meanings of the speakers’ words in a community with their individual dispositions to apply those words. This only follows given certain empirical facts about the idiolects of the members of that population.<sup>1</sup>

---

<sup>1</sup>I thank Douglas Patterson (9/24/09 email) for alerting me to problems in an earlier discussion of this.

We may nevertheless worry about the survival-value of such languages even in empirically fortuitous circumstances. Suppose the dispositions are uniform across all the individuals in such a community—can such individuals *consistently* apply words to objects? Why can't their dispositions be such that at one moment an object is labeled a table, and seconds later it's not labeled as such? Their dispositions *can* be like this, and then, even if they are shared by everyone in the community, the result is disastrous. Whether such idiolects enable *useful* communication, that in addition helps the members of this community survive their interactions with things, turns solely on empirical facts about the dispositions possessed by each member of the community—in particular, the stability of these words across contexts of utterance.

Imagine, for example, that the uniformly-had dispositions to apply words in a community are extremely cautious ones. Perhaps it's a commonly shared disposition of individuals in this community not to apply the word "table" to an object until that object has been examined up close in such-and-such careful ways. Such dispositions do—to some extent—stabilize the application of words across contexts, and they even stabilize (to some extent) the applications of calculational words such as "addition," "multiplication," and so on. Suppose that the members of such a community are disposed to apply a numeral as the result of a computation they've undertaken when and only when those computations have been undertaken and double-checked in such-and-such specific ways (a rushed computation isn't seen as a computation—it's not "finished"). Provided further that everyone has the disposition to give the same answers when computations have been completed (stated from our point of view: provided everyone is disposed to make *exactly* the same mistakes so that they always get *exactly* the same answers), then the idiolects agree, and the applications of the words by the members of this community are stable.

A provisional conclusion is this. At least as far as the discussion has been taken in this section, whether a disposition-meaning language (one, therefore, for which Kripke's three requirements either lapse or are trivially met) is *practically implementable*, turns purely on empirical considerations. It turns on exactly what dispositions the individuals in that community possess, and (importantly) in what ways those dispositions involve uniformities across the population. I haven't yet explored, of course, whether the success of such a language-practice also requires things in the world to possess certain uniformities that correspond in some way to the uniformities in the dispositions of the members of the population. I also haven't yet responded to the challenge, that some philosophers might mount, that—regardless—these aren't cases where the individuals in question have *concepts* that their words can be taken to correspond to, and that they can be taken as applying to objects in the world. Rather (these opponents will say), they just have tendencies to involuntarily express *noises* in ways that are more or less coordinated. The first issue will be dealt with in the course of presenting and discussing the various Robinson Crusoe cases, as I do in the following sections of this chapter and in Chap. 5. The second issue will be addressed in the next section.

### 3.2 Robinson Crusoe in Empirically Favorable Circumstances

Imagine that Robinson Crusoe (hereafter Crusoe 1) lives primarily on coconuts. Hunting for coconuts, however, is arduous (because he has to avoid various large predators). To survive, he has cleverly taught himself to count the number of coconuts he has at a time, how many he eats in a day, and therefore how many days it is before he has to engage in a risky search for more coconuts.<sup>2</sup> He has also studied the territory he forages in, and he has scoped out which trees have more coconuts and which have less—and therefore which trees are better for him to climb. In short, he's in a situation in which his counting skills are vital to his survival.

Included in the description of this Crusoe scenario is the apparently undeniable fact that collections of objects—e.g., the number of coconuts in that tree—have specific cardinalities. It's unsurprising, therefore, that when counting coconuts (and when counting other things, such as the number of days he can avoid hunting for more coconuts) it's important that Crusoe 1 get the numbers he counts *right*. It's important for Crusoe 1 that when he counts the number of coconuts in a tree, that the number he gets corresponds to the number of coconuts actually in that tree.<sup>3</sup>

Crusoe 1's applications of his other words, “coconut,” “tiger,” and so on, must be judged on their adequacy in a similar way. Coconuts are food for Crusoe 1; he, on the other hand, is food for tigers. His word “coconut” should pick out a particular set of items that are food for him; his word “tiger” should pick out a particular set of items that it's best for him to avoid.

This use of “right” is different from the use of it earlier in this book, especially in Chap. 2, where it applied to meanings of individuals that conform to what they meant earlier. What's apparently crucial to this use of “right,” to Crusoe 1's answer being “right,” is that his answer correspond to what's in the tree, so that his actions that result from his count (climbing one tree and not another to shake down its coconuts) yields the number of coconuts he needs.

We can accept, I think, that Crusoe 1 genuinely *understands* the words he has invented: “1,” “2,” etc., “coconut,” “tiger,” and so on. This is because he's *given* disposition meanings to these words: they are to mean exactly what his meaning-urges

---

<sup>2</sup>I'm assuming Crusoe 1 can invent words, and that he can even invent words for complex notions that he acquires—such as cardinalities of collections of objects—on the combined basis of his (innate) dispositions and his experiences. It's doubtful isolated humans can do this, even if some sort of rich innateness hypothesis about human dispositions is nevertheless true. Regardless, the acceptable empirical assumptions (about this) aren't in dispute now. If Kripke's way of deriving objections to the private model of language from the rule-following paradox is right—I'm focusing especially on his second and third requirements for a solution to the rule-following paradox—it doesn't matter how sophisticated we allow a Crusoe's dispositions to be: the rule-following paradox arises in any case.

<sup>3</sup>What's required, that is, is that there is a 1-1 correspondence (between the numerals he applies to collections of coconuts and the collections of coconuts themselves) that preserves the ordering relation of the numerals as he uses them and the ordering relation of “larger than” of collections of coconuts.

(at any time and place) incline him to apply them to. However, given the notion of “right” just utilized, we can ask why so giving words disposition meanings doesn’t lead to a complete disaster. Imagine, however, that Crusoe 1 is in the following rather idyllic situation: (i) his urges to apply words are stable, and (ii) the ways that things appear to him are always the ways that they are. That is, if he thinks he sees a coconut, that’s because he *does* see a coconut. I’ve put these conditions in what some might regard as an illegitimately tendentious way. What, after all, is a coconut to *Crusoe 1*, and relatedly, whatever does it mean to say that his urges to apply words are stable?<sup>4</sup>

So let me put the matter differently. *We* live in a world where sometimes shadows look like coconuts, and where coconuts sometimes look like curled-up cats. Furthermore, it sometimes takes a great deal of investigation to determine whether two things that look alike are in fact alike. That is, the properties that pretty much anything has, coupled with the limitations of our senses, results in our making many mistakes about which things are which. Crusoe 1 doesn’t live in this kind of world. Because of Crusoe 1’s senses, because of the small numbers of kinds of things that there are in his world, and because of the limitations in how these things can differ, everything in his world falls neatly into one or another *recognizable natural kind*. Furthermore, he can identify and distinguish objects at first glance; Crusoe 1 *can’t* have the experience that something looks like one kind of object to him at one time and a different kind of object at a different time. It isn’t possible for him to discover later that things that looked like tigers or coconuts or squirrels now look like different creatures. Finally, he never makes what we would describe as computational mistakes. If he *thinks* there are six coconuts in a tree as a result of counting them, that’s because there *are* six coconuts in that tree.

In such a remarkable world, and with such enviable epistemic powers, Crusoe 1 *can* endow his words with disposition meanings, and enjoy a great deal of worldly success despite this. He *can* understand “7” to mean that cardinal number of a collection of items that he’s disposed to apply “7” to. He *can* mean “coconut” to apply to exactly those things (all identical in their natural-kind-relevant properties) that he’s inclined to apply the word “coconut” to. Because Crusoe 1’s idiolect is a disposition-meaning language, Kripke’s three requirements either lapse or are trivially satisfied. I hasten to add that, because the infinitude requirement lapses, some wouldn’t regard Crusoe 1 as *counting*, as *we* understand counting. Although I’m allowing that Crusoe 1 never makes computational mistakes, I’m not assuming he can count arbitrarily large collections of things. Because there is an upper limit on the size of collections he’s disposed to count, he has only a finite number of number words.<sup>5</sup>

<sup>4</sup>I’ve been describing Crusoe 1 and his world from the perspective of our *own* language. Phrases like, “judging Crusoe 1’s terms on their adequacy,” involve evaluations made in *our* terms. What’s involved in doing this is something that’s explored further in later sections, especially in Sect. 3.3 and in Chaps. 4 and 7. Let’s meanwhile treat this as a preliminary way of speaking of the isolated Crusoe 1 and the issues he raises—a preliminary way of speaking that will later be reevaluated and perhaps corrected.

<sup>5</sup>Because his numbers are finite, I won’t claim he has *our* numerical concepts. On the other hand, it’s not obvious I should deny this either—at least of the numerical concepts he evidently *does*

Nevertheless—or so I argue—Crusoe 1 has a cogent language practice (a cogent idiolect) in which his words mean exactly what he’s disposed to have them mean, and because of extraordinarily nice empirical circumstances, his idiolect is useful to him—indeed it’s indispensable. Furthermore, this is true even though he can’t apply his words wrongly.

Some would say that because he can’t apply his words wrongly he *hasn’t* got a coherent language-practice. (And anyone who would say this would say the same of the community of idiolectal speakers that I considered in Sect. 3.1.) They would deny, in particular, that Crusoe 1’s words correspond to *concepts* that he can be taken to have: cardinality concepts, or the notion of a coconut, etc. Crusoe 1—as I’ve imagined him (so this opponent claims)—doesn’t have (real) concepts because his application of his concepts doesn’t involve his being *guided* by them. Instead, and this is recognized by him, he has urges to make verbal *responses* to things in the world. (The point could be put a little meanly: he’s only got a useful set of verbal *tics*, not a language.)

Three considerations, however, support the idea that his concepts aren’t quite the same as ours (in certain respects) rather than that he doesn’t have concepts at all. First, there is the mental content that accompanies Crusoe 1’s use of his words. Although his understanding of his words explicitly gives those words disposition meanings, that doesn’t stop his words from corresponding to collections of things (in specific circumstances), and to collections that he experiences as similar, or as having certain cardinal properties, and not others. Indeed, this mental content in some cases *constitutes* his dispositions (e.g., he’s disposed to treat such-and-such items as alike because of how he *experiences* them). Second, some of *our* concepts are just like his in similarly having disposition meanings. Consider the ordinary concept of *pain*. Notoriously, this is a concept to which “being wrong” doesn’t seem to apply (at least in the first-person).<sup>6</sup> Lastly, and perhaps most importantly, nothing prevents Crusoe 1 from using *his* concepts in exactly the ways we use *ours*: to describe things in his world, to reason about them, and to decide on a course of action on the basis of that reasoning. Just like us, he can decide to climb one tree rather than another (because of, say, the number of coconuts he takes each tree to have), or he can decide to challenge a group of dangerous predators because there are only two of them, and not three. These striking facts about Crusoe 1’s abilities correspond to similar facts about our concepts and our abilities: those concepts of

---

have: e.g., the ones of collections he has the dispositions to count, and to order in magnitude. I discuss a different aspect of this issue further, momentarily.

<sup>6</sup>Wittgenstein is widely taken to have challenged such concepts in various respects. But his challenges seem to rise directly from rule-following considerations, and so it would be question-begging to press them in this context. (Here I seem in accord with Kripke (1982, 3), when he writes: “The ‘private language argument’ as applied to *sensations* is only a special case of much more general considerations about language previously argued ...” Also, see his footnote 47, 60–61, and the surrounding text footnote 47 is appended to.) I table for future work the other ways that philosophers have thought to undermine this apparent datum about (first person) uses of “pain” and similar words. I discuss these words in Sect. 3.5, and describe how they operate in disposition-meaning languages.

ours to which “correct” or “incorrect” don’t seem to apply, such as “pain,” seem nevertheless to be concepts that can be used together with our other concepts (to which “correct” or “incorrect” do apply) to describe aspects of our world, to reason, and to make decisions about courses of action.

One last observation: It would be a mistake to imagine that Crusoe 1—in such epistemically favorable circumstances—would see himself as omniscient. Certainly he would lack the notion of a “mistake” in one sense of that word: He wouldn’t understand how it’s possible to think something is a coconut when it isn’t. But despite this, he would still understand what it means to be ignorant. He would have experienced (indeed, he would be quite familiar with the experience of) not knowing how many coconuts there are in a tree because he hasn’t counted them yet, or not knowing whether there is a tiger around the next bend because he hasn’t gone round that particular bend yet.

There is a subtlety here that I should at least indicate by distinguishing between two kinds of cases. One is this: Crusoe 1 might face a collection of items—like a pile of sand—that he has no urge to count. He’s neither disposed to apply any of a large number of counting words nor to deny applying them: as far as his idiolect is concerned, there is no answer to the question how many grains of sand there are. He’s not ignorant of the answer (this is how some might want to put it); there isn’t one. But there is a different kind of case, where he feels he would be able to either apply a word or not apply it if he chose to—for example, he may be disposed to deny that “1,” “2,” or “3” apply to the collection of coconuts in a tree. As a result, he knows that there are more than three coconuts in that tree; but without counting (without engaging his dispositions to supply a cardinal number to the collection of coconuts in that tree), he doesn’t know exactly how many coconuts are there.<sup>7</sup>

In the fortunate circumstances our first Robinson Crusoe finds himself—fortunate both because of the idyllically simple world he lives in, and because his senses and verbal dispositions perfectly match what’s in that world—he has no need of the concept of the “correct application” of a word. In this respect his resulting idiolect is different from the language we take ourselves to speak. It is, however, recognizably a language, and recognizably a useful one. This Crusoe thought experiment has therefore already tamed the scope of Kripke’s Wittgensteinian argument against the private model of rule following. Leaving aside the slight terminological abuse of describing this as a case of private rule *following*, the unexpected result is this: In the right empirical circumstances, there are useful private languages because the private rule follower’s dispositions and his world are such that it doesn’t matter that there is no difference, between the isolated individual following a rule and his only thinking that he does.

---

<sup>7</sup>Complications arise if we try to sharpen this distinction because that requires distinguishing Crusoe 1’s dispositions to apply a word from other dispositions he may have that conflict with doing so. It doesn’t matter for the form my argument eventually takes whether this distinction is tenable or not, and so I forego any further discussion of it.



### 3.3 A Robinson Crusoe with Two Sets of Dispositions That He Has Conscious Access to

Let's now imagine a second (slightly) more realistic Crusoe 2 in the same world our previous Crusoe 1 was in—a world of perfectly distinguishable objects that fall into recognizable natural kinds. As before, there are foodstuffs (coconuts, shrimp, etc.) and predators (tigers, boars, etc.); and Crusoe 2's senses, when optimally working, can perfectly distinguish anything from anything else.

As “optimally working” in the previous paragraph intimates, Crusoe 2's dispositions have certain instabilities. Things sometimes seem one way to him, and sometimes they seem another way. Sometimes, when the light is different, or Crusoe 2 is tired, or if he has made a quick decision, or when he has had too much of his home-brewed liquor to drink, what he has an urge to call a “coconut” is something, under later circumstances, he'll have an urge to call “a *yuckonut*.” Sometimes, what he's disposed to describe as a “tiger” is what he's disposed to later call a “shadow” (after staring at it little longer, for example, or after climbing a tree and waiting to see what the thing will do).

A first (and rough) way of putting the difference between Crusoe 2 and Crusoe 1 is that it would be valuable for Crusoe 2 to learn to recognize when he can trust his urges to apply his concepts or his words and when he can't. Kripke (on behalf of Wittgenstein) denies this is a possibility for a Robinson Crusoe. He (1982, 112, footnote 88) writes:

... in the absence of Wittgenstein's sceptical paradox, it would appear that an individual remembers his own 'intentions' and can use one memory of these intentions to correct another mistaken memory. In the presence of the paradox, any such 'naïve' ideas are meaningless. Ultimately, an individual may simply have conflicting brute inclinations, while the upshot of the matter depends on his will alone.

What Kripke (1982, 112, footnote 88) goes on to write is also valuable to take note of:

The situation is not analogous to the case of the community, where distinct individuals have distinct and independent wills, and where, when an individual is accepted into the community, others judge that they can rely on his response ... No corresponding relation between an individual and himself has the same utility.

I disagree. There are two possible cases to consider. The first is the less realistic one where Crusoe 2 can distinguish two internal dispositional states that it's possible for him to be in. The first state induces one set of urges, and the other a different set of urges. For the sake of simplicity, I'll call one state “being rested,” and the other “being tired.” The second, more realistic case, is where Crusoe 2's urges vary over time and place, but he's unable to classify those urges in any neat way, or even recognize (much of the time) which dispositional state he's in and whether his urges have shifted as a result. I take up the second kind of case in Sect. 5.2.



In the first case, Crusoe 2 has a choice between which of *two* possible disposition-meaning languages<sup>8</sup> he takes seriously. That is to say, although Crusoe 2 *speaks* whichever disposition-meaning language he speaks (based on whatever state he's in), he can decide that only one of these languages should be trusted. He recognizes, that is, that the different dispositional states cause him to give his words different disposition meanings, and he can recognize that one set of dispositions (one disposition-meaning language) is superior to the other. He can decide, that is, that one is "trustworthy," and the other isn't. Imagine, in fact, that his dispositions—when he's rested—match the world in the way that Crusoe 1's dispositions match his world. Crusoe 2's dispositions when he's tired, however, deviate. It seems that Crusoe 2 can notice this fairly quickly. When he's tired, his word "coconut" has his tired disposition-meaning of "coconut." But he soon discovers that tired-disposition coconuts aren't all alike in ways that matter to him. Some are good for him; but other ones make him sick. According to the rested-disposition meanings of his words, however, only some of these items are coconuts: the other ones are *yuckonuts*.

The way that Crusoe 2 realizes that his rested dispositional state is trustworthy and that his tired dispositional state is (in some respects) untrustworthy is because he develops and relies on a body of explicit and tacit *gross regularities*.<sup>9</sup> For example, "Eating coconuts nourishes me," and "Eating yuckonuts makes me sick," are two straightforward gross regularities. If Crusoe 2 has the words "eating," "coconuts," "yuckonuts," "nourishes" and "sick," in his rested dispositional language, then he can articulate these gross regularities in his rested dispositional language. He also has these words in his tired dispositional language, let's say, because even when he's tired he can recognize (some) yuckonuts and distinguish them from coconuts. In his tired dispositional language, however, he has a different (weaker) gross regularity: "Some coconuts make me sick and some coconuts nourish me," and perhaps also, "Some yuckonuts make me sick and some yuckonuts nourish me." (This second gross regularity holds if Crusoe 2 also sometimes confuses coconuts with yuckonuts when he's tired.) Being tired or rested, on the other hand, has no impact on his recognition that he's sick or well—eating yuckonuts make him violently ill, let's say, eating coconuts make him comfortable and satisfied.

If he doesn't have the words "nourishes" and "sick," he can still recognize and distinguish between being sick and being nourished: in this case the relevant gross regularities are *tacit* ones. It's reasonable to describe Crusoe 2, in this case, as having the *concepts* of being nourished and being sick, even if he doesn't have the actual words to express the distinctions that he's aware of. Using these concepts, he will grasp the relevant gross regularities as well, even though he can't express them to himself.

Notice that, in this way, Crusoe 2 recognizes that the rested dispositional language is more trustworthy than his tired dispositional language: the trustworthy language

<sup>8</sup>An alternative approach individuates dispositional languages differently so that Crusoe 2 speaks one dispositional language. Nothing turns on this; I would just have to put the forthcoming points in a different way.

<sup>9</sup>I first discuss gross regularities in Azzouni (2000) and again in Azzouni (2010b).

has more precise gross regularities that enable him to more successfully predict the outcomes of his actions. Similar gross regularities—superior rested dispositional ones and inferior tired dispositional ones—can be formulated using his number concepts. Because of this, Crusoe 2 can recognize, when he’s running out of coconuts (and needs to get more the next day) that he should get enough sleep so that he’s using his rested language the next day and not his tired language. (In general, this too is tacit knowledge—something he recognizes but can’t put explicitly into words.) He doesn’t want the experience of gathering a lot of produce, and after a good night’s sleep, having to throw a lot of it away, or worse, getting violently ill because he ate some of it while he was still tired—bad experiences he’s had in the past.

### 3.4 What Does “Better Matching the World” Mean to Crusoe 2?

I wrote this in the last section: “Imagine, in fact, that [Crusoe 2’s] dispositions—when he’s rested—match the world in the way that Crusoe 1’s dispositions match his world.”

Talk of “matching the world” is metaphorical talk (that certain philosophers nevertheless take very seriously). But can this kind of talk mean anything to Crusoe 2? Can we say that, in recognizing that his rested dispositional language enables him more success in the world than his tired dispositional language, that he realizes this because the terms of his rested dispositional language better match the world than the terms of his tired dispositional language do?

In asking this, I’m not asking the practical question of how Crusoe 2 manages to recognize which language is better. It’s relatively obvious how he manages that because it’s clear, and clear to him, how one set of dispositions keeps getting him in trouble, and the other set doesn’t. We might be tempted to spell out the matching suggestion this way: Crusoe 2 *thought* something was a coconut, but after he ate it, he got violently ill, and that’s when he realized it wasn’t a coconut. But this is wrong. To see why, let’s use ordinary quotes to describe words in the language *we* use to describe Crusoe’s situation; let’s use r-quotes for his rested language, and t-quotes for his tired language. We can’t say he thought the item was a “coconut,” but then he discovered it wasn’t. That uses *our* word, not his word. He, however, *didn’t* discover the item wasn’t a ‘coconut’, because it *is* a ‘coconut’. We *can* say he discovered it wasn’t a ‘coconut’. But it’s not that he *thought* it was a ‘coconut’. He wasn’t using his rested-disposition language when he uttered the sound, kō’k(schwa) nut’; he was uttering his word ‘coconut’, which applies correctly to what he applied it to. (And this is true even after he gets violently ill—as long as he’s still tired.)

The trouble that Crusoe 2 has gotten into, though, seems to be something that can be expressed in either language that’s available to him. According to how Crusoe 2 can speak in his tired-disposition language, the trouble is that some ‘coconuts’ make him violently ill and some don’t. But this doesn’t imply that they’re not ‘coconuts’.

On the other hand, all 'yuckonuts' make him ill. Either way, he ate something he shouldn't have.

*We're* tempted to say: The rested-disposition language *really* describes coconuts and yuckonuts *as they are*. After all, Crusoe 2's word 'coconut' agrees in its extension with *our* word "coconut," and our word gets it right. But what's wrong with the extension of 'coconut'? *We* think some 'coconuts' are coconuts and other 'coconuts' are yuckonuts.<sup>10</sup> But so what? The words in each of Crusoe 2's meaning-disposition language have different scopes—they range over different collections. And there is nothing wrong with *this*.

*We're* tempted to say in response: The set of collections of objects as categorized by Crusoe 2's rested-disposition language accords (better) with the real resemblances among the objects so collected together; and the set of collections of objects as categorized by his tired-disposition language doesn't. More is tempting: The rested-disposition language *matches* its words to the natural kinds that are in the world, and the other doesn't. Coconuts and yuckonuts simply don't belong together in a kind.

There are two issues here. The first is: what gives *us* the right to make this claim? This will be analyzed later—see especially Chap. 4. The second issue is that, in any case, Crusoe 2 hasn't access to these considerations. He's got *two* phrases 'real resemblance' and 'real resemblance'. That is, each of his languages—by its own lights—characterizes the real resemblances in its own way. He also has two phrases that correspond to our words "natural kinds." Each of his languages—by its own lights—characterizes the "natural kinds" differently. With what words is Crusoe 2 supposed to make sense of the idea that the words in one disposition-meaning language are "more accurate" to how the world is than the ones in the other disposition-meaning language?

Notice the point. It isn't (or it isn't *yet*) that *we* can't make sense of the idea that one language better fits the world than the other does. As I've been indicating, *we* apparently do exactly that by comparing the extensions of his words to *ours*. The issue at the moment is this: how is *Crusoe 2* supposed to make this distinction—for himself—have content that goes beyond the already stated fact that one language enables him to navigate the world better than the other one does because one language yields sharper gross regularities than the other one does?

Let's table the question of *our* right to describe Crusoe's words as fitting (or not fitting) the world by imagining that, from *outside* both of Crusoe 2's possible meaning-disposition languages, God has a language each word of which, because of *His* powers, refers to a collection that contains only the members of perfectly natural kinds; imagine too that *His* phrase "natural kind" refers only to the collection of those collections. (Don't ask how *God* makes *His* language do this; He's *God*, after all, and a supernatural ability like this one comes with the territory.) Using *His* own

---

<sup>10</sup> Maybe the example has to be spelled out a little. Some coconut trees are infected by a parasitical vine that grows in and on the coconut tree, and has its own nuts. Crusoe 2, when rested, can distinguish the two kinds of nuts. *We* recognize that the yuckonut plant isn't even a tree, although Crusoe 2 doesn't know *that*.

language, God can describe Crusoe 2’s situation this way: There *are* the ways that things genuinely resemble one another (or don’t), and so they really belong to the same natural kinds. The ways that things really resemble each other *affect* Crusoe 2. One of *Crusoe 2*’s languages—the rested meaning-disposition language—matches the world’s resembling things in these respects; the other doesn’t; and so when Crusoe 2 chooses the rested meaning-disposition language, he can navigate what’s in the world more successfully than he can if he chooses the tired meaning-disposition language. In the first case he categorizes and groups things by the ways that they genuinely resemble each other, and in the second case he fails to do this.

This way of putting things (*God’s* way of putting things), as noted, isn’t available to *Crusoe 2*. Despite this, *Crusoe 2*’s choosing the rested meaning-disposition language over the tired meaning-disposition language is rational because he can recognize the superior choice by how it improves his ability to make his way in his world. He sees that he’s enabled to succeed when adopting one language in ways that he isn’t so enabled when he adopts the other language. He can’t capture this difference, as God can, by comparing the two languages directly to the world, and seeing that one “fits” the items in the world better. For what can he *mean* by this use of “fit”?

Notice what Crusoe 2’s problem is. It’s that he has—let’s say—the ability to understand that one language is superior to another because it enables his success in the world. But *this* is not a necessary and sufficient condition for a language better “fitting” the world. He has, further, the ability to imagine that one language will be superior to all other possible languages (for all time), and in every way, for enabling him to succeed in the world. But this too isn’t a necessary and sufficient condition for a language “fitting” the world. I’m suggesting it isn’t possible for Crusoe 2 to understand what a language fitting the world can mean. (And, I’m also suggesting Crusoe 2’s problem with understanding how God can make His language fit the world is deeper than it initially appears. Perhaps, *we* don’t understand what this can mean either—this is a question I dedicate Sects. 7.2 and 7.3 to probing further.)

Why can’t Crusoe 2 understand the idea of a language fitting the world? The reason is this. By assumption, his languages are disposition-meaning languages; relatedly, his concepts are disposition-meaning concepts. That means we’re to describe his understanding of any concept in terms of his dispositions to apply that concept. But it’s precisely this that goes missing in the case of his view that one language is superior to another by virtue of its better “fit” with the world. That’s not how he responds to the comparison of two disposition-meaning languages. *He* responds only in terms of which language enables him to better succeed in the world, and that isn’t the same thing as the words of such a language better fitting the world.

One thing Crusoe 2 can *try* to do is officially adopt one language, and deny that the tired disposition meanings determine the meanings of *any* of his words. We might think the way he should indicate this is to say, using the rested meaning-disposition language: “I was tired, so I *thought* I was carrying a coconut, but I wasn’t. I *made a mistake*. It was a yuckonut.”

This won’t work. The problem is that both his languages are *disposition-meaning* languages. He doesn’t use the terms of either of these languages unless he’s *disposed* to use them. When he’s tired, he speaks his tired-dispositional language; if he

denies his tired dispositions determine the meanings of the words in his tired dispositional language then those words (when he's tired) are *meaningless*. For this reason, he can't say: "I was tired, so I thought I was carrying a coconut"—and mean "coconut" in the rested sense. After all, when Crusoe 2 is tired, he's not using *that* term at all; so in what sense can he "think" he's carrying a 'coconut' when he's tired?

Notice this is true, even if we accept that Crusoe 2 knows (or remembers) what the terms in his other language refer to. He can say, "If I was rested, I would have thought that was a 'yuckonut'." He's right about this; but his being right doesn't make the formulation, "I thought I was carrying a 'coconut'" correct.

Because Crusoe 2 is aware that there are two meaning-disposition *languages* that he speaks at different times, and because he knows he has less success with one than with the other, he *can* describe the situation metalinguistically. He can say instead: I used a less useful word 'coconut' because I was tired. In doing so, he still describes himself as having made a mistake; but a mistake about the word he used, and not a mistake about something in the *world*. Or, equivalently, he can say, "I had the wrong dispositions when I uttered the sound, kō'k(schwa) nut' because I was tired." In the second case, he thinks of himself, because he was tired, as mistakenly employing the wrong dispositions for the single sound kō'k(schwa) nut'. In the first case, he instead thinks in terms of there being two different words, individuated by different dispositions, and he has used the wrong word. For Crusoe 2, the difference seems to be merely terminological.

To repeat, Crusoe 2 can choose one language over the other only because it's superior in enabling him to navigate his world—and not because (as God can) he can see that one "fits" the world in a way that the other doesn't. His doing badly if he adopts one language, and his doing better if he adopts the other, is an objective fact. But it's not a fact *he* can express by a comparison of how languages purport to contour the world to how the world itself is contoured.

We've returned to the concept of *fit*. Some philosophers might think that Crusoe 2 can think of *fittingness* as an *explanation* for why one language is superior to the other: the superior language fits the world better—*that's* why it's better for navigating his world. Crusoe 2—so some of these philosophers might say—is engaging in an inference to the best explanation for his success in the world: his success must be due to the fact that his terms are picking out things in the world in the way that those things are. His success must be due to the fact that the ways that his terms carve the world matches the way that the world is carved—at its "joints," as it were. He doesn't *see* that this is so; he *infers* from his success that it must be so.

We still have the problem we had earlier: what can Crusoe 2 *mean* by these metaphors: "fits better," "carves the world at its joints" when he gives this kind of explanation? (How can he *understand* these metaphors?) How does what he means by these phrases go beyond the evident fact that he can better navigate the world by means of the language he has chosen? One thing these metaphors *can't* mean is this: Given the "correct" description of the world, the superior language describes the world in the same way. It can't mean this because Crusoe 2's notion of a "correct description" of the world can only be a description that's a better *fit* than any other description (it's

one that succeeds in carving the world at its joints). And now it's clear that the same problem has arisen again: what are these metaphors supposed to mean to him?

Think of a map. Some maps, we think we can say, “perfectly” describe the terrains they characterize. But how does this go beyond: using the map can't mislead us in some way? We think it does go beyond this because we can imagine, after all, a piece of paper with lines on it, and we can also imagine a terrain the contours of which the lines in the map fit perfectly. (The lines on the map are isomorphic to the contours of the terrain *modulo* the respective sizes of the map and terrain.) This *imagery*, however, faces exactly the same problem. Our “image” of the terrain is still a description of the terrain “without language or concepts,” one that we have abstracted away from the input of our (visual) dispositions; and we're taking ourselves to recognize that the map fits this supposedly wordless (and nonconceptual) characterization of the terrain. Considerations like these suggest that Crusoe 2's problem is our problem.<sup>11</sup>

Suppose that Crusoe 2 is quounting, rather than counting. As we've seen, *God can say*: Crusoe 2 will do badly because he will treat collections as identical in cardinality when it matters that they're not.<sup>12</sup> On the other hand, imagine there is someone else on the island with Crusoe 2 (*Friday*, say), and suppose that Friday and Crusoe 2 both have exactly the same rested, and tired, dispositions. If Friday (wisely) chooses to negotiate with Crusoe 2 only when Friday's rested, and speaking his rested disposition-meaning language, but Crusoe 2 instead (foolishly) negotiates with Friday when Crusoe 2's speaking his tired disposition-meaning language, then Friday can take advantage of Crusoe 2 in *all* sorts of ways. (Friday will realize that he can use Crusoe 2 as a *money pump*—as it's commonly put.) Suppose Friday describes this from the vantage point of his own rested language (because he's currently speaking it). He might naturally say that because Crusoe 2 *thinks* he has 5 coconuts, he's willing to trade them for Friday's 5 sardines (but, according to Friday, Crusoe 2 is actually trading 57 coconuts for 5 sardines).<sup>13</sup>

Friday's authority for his claim is only that he can successfully take advantage of Crusoe 2. He hasn't God's reasons for his claim (even though, let's say, his and God's languages are word-for-word translatable in the sense that Friday's application of his words matches God's application of corresponding numerical terms).

---

<sup>11</sup> Many philosophers over the years have claimed—against the correspondence theory of truth, for example—that one's words can't be compared to reality. It's not always obvious what such philosophers mean when they say this. One thing some of them might have meant is this.

<sup>12</sup> It looks like God can say this only if He doesn't “translate” (homophonically correlate) Crusoe 2's terms to his own. This is for the same reason that a rested Crusoe 2 can't say of himself that when he was tired he *thought* he was carrying a ‘coconut’ and why we can't say that Crusoe 2 *thought* he had a coconut when he had a yuckonut. God *can* say that Crusoe 2's tired dispositional language makes him treat coconuts as yuckonuts, or distinct cardinalities as the same cardinalities, without describing what Crusoe 2 is *thinking* when this happens. God can also say that Crusoe 2's words aren't picking out real differences and similarities among objects—differences and similarities that bear on his (Crusoe 2's) well-being—and that's why he'll do badly. This issue recurs in Sects. 3.5 and 6.4.

<sup>13</sup> Again there is the issue of footnote 12. Friday can say this by “translating” the terms of Crusoe 2's tired dispositional language to his own; but that's to misdescribe Crusoe 2's thinking.



God knows what He knows, by assumption, because He makes His words fit the world metaphysically—or, perhaps because He made the world to fit His words. Friday hasn't got God's supernatural powers: so Friday can only know that he's doing far better than Crusoe 2 is.

It might be thought that the opening description of Crusoe 1 and his world in Sect. 3.2 is a description of him from a "God's eye point of view." Perhaps it's better not to attribute such powers to *us* any more than to Friday. Perhaps it's better to say that we describe Crusoe 1 from a "Friday's point of view." Then our talk of Crusoe 1's dispositions perfectly fitting his environment means this: If we were on the island with Crusoe 1, we wouldn't be able to make this Crusoe 1 into a money pump: we wouldn't be able to exploit him.

Let's return to the isolated Crusoe 2 evaluating his two meaning-disposition languages. To repeat: the evident fact that Crusoe 2 does better with one language than with another is what enables him to evaluate them. And, having chosen, he can now express his recognition of one language's superiority, as we saw earlier, either by faulting his dispositions that his use of words shouldn't rely on, or by faulting the language (with particular disposition meanings) that he has inadvertently slipped into. Doing so, however, doesn't (and can't) rely on Crusoe 2 knowing the metaphysical correspondence fact that terms in the superior language carve reality in a superior way—e.g., more accurately—than terms in his earlier language did.

Nevertheless, the foregoing suggests that in the right circumstances, Crusoe 2 can realize that it's appropriate for him to trust *some* of his dispositions, and not other ones. Necessary to this ability, of course, is that he can identify (some of) his internal states. *Given* that ability, his favoring some of his dispositions over other ones *isn't* an arbitrary matter that depends on his will alone (*pace* Kripke's Wittgenstein). Rather, it turns on objective facts about the success value of the different disposition-meaning languages that he uses (or tries to avoid using). "Success," so understood, turns on Crusoe 2 being able to evaluate his own well-being. I turn to exploring this in the next section.

### 3.5 Introspection

I've suggested that Crusoe 2 distinguishes his dispositions on the basis of the subsequent success he experiences when relying on them. Crusoe 2, so I'm assuming, can distinguish and grade his well-being subsequent to events that affect it. I'm also assuming he can recognize how his well-being is affected by those events. It's not necessary to assume Crusoe 2 is entirely objective about this, or that he's always right. It *is* necessary, however, to assume that he's pretty good at recognizing when what he has eaten has made him ill or when an animal is dangerous, or has hurt him, and so on. In allowing that he can do all of this, I'm, of course, also assuming that Crusoe 2 has a number of dispositions that enable him to learn from his experiences.

One issue that these assumptions raise—right at the start—is that different Crusoes can have different “systems of values.” The Crusoes in this book are straightforward simple creatures with values we find easy to understand: they want “creature comforts,” they want to survive, they want to avoid pain and discomfort; they want meals that don’t make them sick. But other Crusoes are possible: ones who disdain creature comforts, who wish (masochistically) to be eaten by tigers, who embrace pain and discomfort and death. The appropriate notion of “success” seems relative to the particular values of the Crusoe in question.<sup>14</sup> Consider again the case where Crusoe speaks two languages: the tired disposition-meaning language and the rested disposition-meaning language. Surely it’s the case that subsequent success for a Crusoe with certain (self-disaffirming) values turns on his making sure (as much as possible) that he execute tasks when speaking the tired disposition-meaning language rather than the rested one. For in that case, he increases the likelihood of his own death—which, let’s say, he desires to occur in a certain way (for example, by an accident brought on because of his own incompetence).

There is no denying the relativity of a Crusoe’s success to what that Crusoe wants. “Success” must be, therefore, at least partially relativized to the aims and desires of the Crusoe in question.<sup>15</sup> Two points, however, have to be made about this. The first is that one aim of this book is to introduce a series of Crusoes with a broad range of dispositions, in a family of environments, where they can engage in cogent private rule-following. I intend, ultimately in Chap. 5, to root that cogency in a trajectory of modifications of a Crusoe’s own dispositions that lead to “successes” that the Crusoe can be aware of, or at least that he can respond to, and that—at the same time—are objective. I’m *not* claiming we can attribute to such a Crusoe *any* set of dispositions whatsoever, and expect him to be able to engage in cogent rule-following. That’s not true. So in this book I stick to relatively simple cases: ones where a Crusoe wants to survive, he wants creature comforts, etc., and I’m going to defer “unhealthy” Crusoes for other work.<sup>16</sup>

The second point is this. Although “success” must be relativized to a Crusoe—more specifically, to his values, still more specifically, to what he wants—at the same time this success must be objective. It is, that is, an objective fact whether Crusoe really is better off as a result (relative to the standards provided by his values) when he takes there to be so-and-so many coconuts and not a different number,

---

<sup>14</sup> My thanks to Mitch Green (oral communication—August 8, 2009) for raising this issue, and prompting the paragraphs to follow.

<sup>15</sup> “Partially,” because there are complications. We sometimes allow our evaluation of someone’s success to include an evaluation of their values. Thus, we may deem someone a failure because (although given their values, they have succeeded) we regard the successful satisfaction of *those* values a failure. These complications don’t directly bear on the role of “success” as it’s used here.

<sup>16</sup> I don’t mean to suggest, by “unhealthy,” that such Crusoes have values that we should ignore or rule out. Nor do I want to claim that such Crusoes couldn’t engage in cogent private rule-following despite their “unhealthy” values. Many clearly could. Rather, I’m pursuing straightforward cases of Crusoes (with, that is, fairly straightforward dispositions) to simplify discussion. I aim to demonstrate private rule following is possible under a wide range of circumstances, so this restriction is acceptable.



and so on. It's an objective fact whether he's better off grouping the creatures in his world this way rather than that way. These objective facts are relative to what it is that Crusoe wants; but they are objective nevertheless.

This answer raises two further issues. The first is that such objective facts, about whether Crusoe is better off or not, seem ones that must be expressed in a language that really describes the world as it is. At the same time, it also seems that such objective facts must be rooted in the propositional attitudes of Crusoe: that *he* wants such-and-such, and that *he* doesn't want so-and-so.<sup>17</sup> But any such description of Crusoe's propositional attitudes must be—so it seems—couched in some language or other. Since Crusoe articulates to himself that he wants a coconut, and that he wants to avoid a tiger, it seems that such attitudes must be couched in whatever language he (Crusoe) is using—in particular, *his* words “coconut,” “tiger,” etc., seem needed. But his language doesn't seem to allow a description of the appropriate failures. After all consider Crusoe 2 using the tired disposition-meaning language: he *wanted* a 'coconut'. And he got a 'coconut'. Where's the failure? If we describe such attitudes, using an objective language (God's language, say), then we are no longer expressing Crusoe 2's propositional attitudes as he would express them, and perhaps not even as he would recognize them.

The first move to make is to deny that Crusoe 2's success is to be couched in terms of “coconuts” and “tigers.” Rather, it's to be couched in terms of the satisfaction of his more primitive desires—hunger, fear, safety, and so on. It's in terms of these concepts—by tacit (or explicit) gross regularities containing them—and not with 'coconut' or 'coconut' alone that his successes and failures are to be indicated.

But this seems to just push the problem back a step. There are two options. The first is that Crusoe 2 has words for these psychological states as well: 'hunger', 'hunger', 'fear', 'fear', 'comfort', 'comfort', and so on. And so, as my r- and t-quotes indicate, it seems that the same issue arises for these words that arises for his coconut- and tiger-words. The second option is for Crusoe 2 not to have words for these psychological states. But it's unclear why this second option would enable an escape from the concern about the language of Crusoe 2's propositional attitudes. For in any case, he certainly can *recognize* that he's hungry or fearful. And when he's tired, he might think he's hungry when he's not, fearful when he's not, or in pain when he isn't. So it looks like we still have the same problem because even without explicit words, Crusoe 2 is still saddled with two sets of concepts.

This problem can be neatly solved if we attribute to Crusoe 2 a certain introspective access to his own states—so that he usually doesn't think he's hungry when he doesn't feel hungry, fearful when he isn't, or in pain when he isn't. In doing so, Crusoe 2's psychological self-impressions are linked with *certain* of his psychological states. Crusoe 2, at least in these respects, is always correct about certain of his own psychological states; and that means that if his success and failure is couched in terms of *these* concepts, pain, discomfort, feeling hungry, and so on (whether he articulates them or not), the problem is solved. For Crusoe 2's propositional attitudes, his desires and hopes, are objective when couched—not in terms of coconuts

<sup>17</sup>I owe my attention to this concern to Stephen Schiffer (oral communication—August 8, 2009).

or tigers but—in terms of pain, discomfort, and hunger: these very notions could be used by God if He described Crusoe 2's propositional attitudes. Furthermore, these concepts (or the words that indicate them) are rigid: they don't shift as Crusoe shifts in his languages.

This solution doesn't pose the danger of creating an incoherence in the Crusoe 2 thought experiment, or in its descendants; it also poses no danger of showing that Crusoe 2, counter to my claim, fails to engage in cogent private rule-following. Rather, the dangers it poses are two. First, it may be that in positing Crusoes with such access to their internal states, we're deviating from the assumption that they speak disposition-meaning languages. Instead of them having dispositions to apply "hunger," for example, we seem to be stipulating that they apply "hungry" to themselves exactly when they are hungry. But the dispositions behind the hunger-urge itself and the meaning-disposition to apply the word or concept "hungry" to oneself are in principle different. It shouldn't be simply presumed that when one is hungry, one has the urge to apply "hungry" to oneself. More generally, being *in* a state, and having *words* for those states are two different things; the state itself may be (or induce) dispositions, but it doesn't follow that these are (or are even related to) dispositions to apply words. The second danger is that this solution may make Crusoe 2 (and the descendant Crusoe cases discussed in later chapters) so artificial that the cases of private rule following they allow are too *recherché* to take seriously. I address these concerns in the rest of this section.

Let's characterize Crusoe 2 as being in certain states: being tired, being rested, being hungry, being in pain. And let's also characterize Crusoe 2 as having certain concepts (or words): "hungry," "in pain," and so on. Is it plausible that Crusoe 2 has the meaning-disposition to apply "pain," "hunger," to himself *when and only when* he is in pain, is hungry, and so on? No—that's too strong; and it's not needed. We *did* presume something analogous for Crusoe 2's rested dispositions with respect to coconuts and the like. But let's be more realistic with his introspective concepts and words. It certainly *isn't* required that Crusoe 2 (when rested, say) has terms, 'pain', 'fear', and so on, that apply to his introspective states when and only when God would apply *His* terms to Crusoe 2's states. Perhaps these terms deviate from God's usages once in a while, or every so often. (We know that someone can be affected by certain diseases so that their uses of terms, "thirsty," for example, deviate widely from God's application of these words.<sup>18</sup>) And perhaps, Crusoe 2's terms, 'pain', 'fear', and so on—his tired disposition-meaning language terms—deviate even more from God's usages. In this case, we must allow that he has two sets of introspective concepts (or words), for example, 'fear<sup>t</sup>' and 'fear<sup>r</sup>'.

If these paired concepts deviate too far from one another (and from his internal psychological states—as God sees them), a Crusoe 2's capacity to succeed in the world will collapse. But as long as these words *largely* agree in their extensions, correspond enough with his internal states, and thus largely agree with God's words, they will provide Crusoe 2 with the ability to recognize success and failure in his actions (relative to his values), and consequently to be able to recognize superior

<sup>18</sup>Descartes (1993, 58) speaks about "dropsy" in this respect.

disposition-meaning languages that enhance that success. Furthermore, when describing his wants, outsiders (God, ourselves) will be able—with only a little inaccuracy—to use their (His, our) own words “fear,” “hunger,” and so on, to describe Crusoe 2’s relevant states.

It’s worth noting that more or less successful correlations of one’s self-descriptions with one’s own creature-comfort levels seem to fit very well with our self-ascribed psychology. Even though contemporary pop-psychological literature is saturated with fallout from studies that show how deluded we can be about our own motivations, and how poorly we often judge our capacities and our performances, it’s still true that we seem—broadly speaking—to be aware of when we are in pain, or in discomfort. Indeed, we’re pretty good, when it comes to basic creature comforts, at recognizing when we’re better off (or worse off) than we were before. I stress again: our skills in this area don’t have to be all that good to enable successful private rule following. More significantly, our varying abilities in this respect translate rather directly (leaving aside the factor of luck) into how well we can engage in private rule following. Not only is that all that’s needed; it’s the best one should expect.

I’ve fallen into speaking in terms of our getting our introspective states *right*; and facilitating this way of speaking (as I sometimes do) by alluding to a God’s eye view. But the situation, with respect to Crusoe 2’s terms describing his inner states, and those terms used by others (God, us) to describe those states is exactly the same as the situation with respect to Crusoe 2’s terms for the objects in his world (coconuts, yuckonuts) and those terms used by others (God, us) to describe those objects. In particular, we can redescribe our “objective” way of comparing Crusoe 2’s terms for his introspective states to our terms for his states, by seeing what we do from a Friday perspective instead. Instead of speaking of better and worse correlations between Crusoe 2’s self-attributions of his internal states and those states themselves, we can instead describe the matter in terms of his capacity to *successfully* evaluate his decisions in terms of their successful or less successful outcomes. To be in pain or hungry or scared is to be (relative to a set of values) worse off than not to be in pain or hungry or scared. And to have words that apply to one’s own states (that allow one to navigate one’s own states) so that (all things being equal) one minimizes being in pain or hungry or scared induces a more successful way of navigating the world than otherwise. Generally speaking (and in the long run), creatures of the latter sort do better than creatures of the former sort.

The forgoing considerations are enough to blunt the worry about the private rule-following cases like that of Crusoe 2 being too *recherché*. Two additional points about this: First, these points about introspection also apply to Crusoe 2’s memories. This is important because his memory is crucial to his being able to comparatively evaluate his languages; he has to *remember* his earlier psychological history. This requires a battery of concepts (or words) that describe his introspective past. I took on, without comment, the implicit idealization employed by Kripke (and mentioned in Chap. 2) that the subject have perfect recall of her past psychological states. But this assumption of Kripke’s can’t be blithely adopted for speakers of disposition-meaning languages. Psychological terms, even when self-applied, must be given disposition meanings. I’ve already suggested that doing so, and allowing these meanings

to accord fairly closely to the actual states of a subject (as God would describe them)—in the case of hunger, fear, and the like—isn't psychologically unrealistic. Memory may seem different—especially given the recent psychological literature about the inaccuracy of memory.<sup>19</sup> Again, an exact correlation *isn't* required. It isn't required that Crusoe 2 is disposed to think he remembers A when and only when A has happened. I'm only requiring enough of a correlation between Crusoe 2's two languages—with respect to introspective states and with God's—for a Crusoe 2 to be able to evaluate the results of the two languages. To speak as God might, a Crusoe 2 who tends to be systematically deluded about what he remembers having happened with 'coconuts' when he's well rested will not survive. It's an empirical matter, how tight a correlation is needed between what Crusoe 2 remembers and what happens. For the purposes of this series of thought experiments, we posit it as good enough for him to make the needed comparisons between his two languages.

Crusoe 2, perhaps, may strike some as quite unnatural, given his introspective terminology. For Crusoe 2 knows himself to be speaking dispositional languages. Perhaps it's easy to understand that he thinks a 'coconut' is whatever he's disposed to call a "coconut" when he's rested; but can it really make sense for him to think that a 'memory of eating a coconut yesterday' is whatever he's disposed to (right now) call a "memory of eating a coconut yesterday"? This way of thinking about our introspective states, nevertheless, isn't alien to us. In particular (and despite what Moore claims about "remember"<sup>20</sup>) we do often say things like, "I remember putting my wallet back in my coat, although I obviously didn't." We also talk about thinking we remember such and such, or thinking we're hungry or well-rested, when it's clear we can be wrong; all that's driving us to apply these phrases "thinking I remember" or "thinking I'm thirsty" is that we're *disposed* to apply these phrases. This is evidence that phrases like "thinking I'm thirsty," that is, *are* meaning-disposition terms. After all, when speaking this way, it's very hard to make sense of being wrong about this.

### 3.6 Last Remarks

The cogency of certain notions (for example, that one is "justified" or "not justified" in applying a word the way one does, or that it has been applied "correctly") have lapsed for speakers of disposition-meaning languages like those Crusoe 1 and Crusoe 2 speak—just as Kripke's analysis of the rule-following paradox predicts. Nevertheless, disposition-meaning languages, even in the simple forms given in this chapter, are successful "private languages" that are pretty useful for their speakers. I've shown, furthermore, that such languages can be *objectively* compared by speakers who speak them, and compared by speakers *all by themselves*. It can be determined objectively by the speakers of those languages whether *those* speakers are

<sup>19</sup> See Baddeley et al. (2015), for a textbook discussion of much of this literature.

<sup>20</sup> Moore (1962, 210–211).

better or worse (using one or the other language) in enabling navigation among the things in the world. And this is the case despite such speakers being unable to make sense of notions like, their words “fitting the world,” or “carving the world at its joints.” These are surprising results.

Despite my qualifications about memory and access to internal states (in the last section), Crusoe 1 and Crusoe 2 remain, of course, psychologically unrealistic. The main unrealistic bit, for the purposes of this book, is that the relationship of their languages to their dispositions are transparently visible to them (ours generally aren’t). In Chap. 5, I return to designing Crusoe thought-experiments, but I endow those Crusoes with progressively more realistic psychologies. In the meantime (that is, in Chap. 4), I turn to blunting a family of natural challenges to my claim, specifically in Sect. 3.4, that “carving joints” or “fitting the world” notions have lapsed in the context of Crusoe 1 and Crusoe 2. These are challenges that originate in the work of David Lewis.

## Chapter 4

# Reference Magnetism

**Abstract** The influential response of David Lewis to the rule-following problem (posthumously described as “reference magnetism”) is described. Three distinct approaches are traced to Lewis’s seminal papers. The first treats the world’s structure as metaphysically providing resources that supplement what individuals have to determine reference. Our words (concepts) have determinate reference beyond the psychological and neurophysiological resources of any individual, and beyond what any community of such individuals can manage on its own. The second treats interpreters of natural languages as required by semantic theory (along with background scientific practice) to impose determinate reference on the terms of those languages, along with natural kinds as the relata of the kind terms of those languages. The third treats the *a priori* constitutive imposition of natural kinds to be required by the Moorean facts of determinate reference and by semantic theories (that presuppose determinate reference) being “the only game in town.” I show that none of these approaches to reference magnetism work.

### 4.1 A Second Family of Straight Solutions to the Rule-Following Problem: Reference-Magnetism Approaches

Straight solutions to the rule-following paradox try to preserve classical correspondence metaphysics between our assertions and the world. This can be by attempting to show that Kripke’s meaning sceptic hasn’t noticed resources we actually have for making our references to things in the world determinate, or resources we have for fixing definitively that we intend to count instead of to quount. Or it can be by supplementing in some way an individual’s otherwise insufficient capacities to manage this by resources located outside that individual. Kripke’s meaning sceptic claims that the needed resources (the grounding facts) aren’t to be found in an individual’s psychology or neurophysiology. One family of straight solutions—that I criticized in Sect. 2.5—respond to this challenge by locating the additional needed resources in the community of speakers that the individual belongs to, or in a community of speakers that an *interpreter* of that individual belongs to.

A different family of straight solutions—ones associated most prominently with the work of David Lewis—locates those resources either in the world itself, or

(depending on the interpretation of Lewis's approach) in how we best theorize about the world and ourselves. The claim is that something about the world itself—issuing directly from correspondence metaphysics or from scientific practice—provides additional resources (grounding facts) that deflect rule-following scepticism. This second family of straight solutions (reference-magnetism approaches) is certainly as popular a response to the rule-following paradox as are the versions of the sociological response that Kripke attributes to Wittgenstein.

I turn, in this chapter, to showing why none of the reference-magnetism approaches work. I show, in later chapters, why they aren't needed. One point: the metaphysics literature that takes the Lewis approach seriously has only scratched the surface of the textual complexity of Lewis's presentation of his position. The problem is that Lewis's prose style (like that of other good philosophical stylists—Hume, in particular) masks complexities and ambiguities in content. As you'll see in what follows, I extract three distinct positions—differing strikingly in their assumptions—from the relevant Lewis discussions (Lewis 1983, 1984). Lewis's rhetoric and his arguments exhibit some ambiguity among all three positions—although I briefly argue that a specific one can be pinned on him. The important point is that none of the positions successfully respond to Kripke's rule-following sceptic; for the same reason, they don't handle "Putnam's Paradox" either—although I won't dwell on this latter point. Since all three responses collapse because of internal difficulties, they can't be used against *any* "paradox."

## 4.2 Can Crusoe Use the Way Things Are in the World as Standards for His Words?

Some philosophers will challenge my claim in Sect. 3.4 that Crusoe 2 can't compare the "fitting world" virtues of his tired and rested languages. Surely, they will claim, Crusoe 2 *is* comparing these two disposition-meaning languages—and not by asking himself which set of dispositions enables him to better navigate his way around the world, but by the very different question: Which set of words (best) captures what's out there? I've already argued that this *isn't* possible for Crusoe 2, but the opposing view that somehow this *is* possible is so seductive (and so popular) that it deserves a chapter-wide discussion of its own.

It's certainly natural to think that Crusoe 2 recognizes that he wants his words to refer to collections of things in the world that *are* alike, and not merely to collections of things that he's *disposed* to treat as alike. But how is he to *understand* his words as doing this? More strongly, how does he *make* his words do this? In Sect. 3.4 I entertained the idea that God has a language where each word refers to a natural kind, and His phrase "natural kind," as well as phrases like "the same kind as," and so on, pick out, and operate with respect to, all and only the natural kinds that there are. We don't need to explain *how* God's language has this property; He's omnipotent (and so we can stipulate this ability of Him without—and this is



extremely convenient—*understanding* how it's possible). We *do*, however, need to explain how a Robinson Crusoe can do anything similar.

So now we're engaging in a thought experiment with yet a third Crusoe, Crusoe 3. Imagine his island world is like that of Crusoe 1; imagine also that his senses aren't like Crusoe 1's and Crusoe 2's (when well-rested) but more like ours—pretty good, let's say, but liable to deceive him under a wide range of circumstances. "Liable to deceive" him means that he'll classify items together (as coconuts or tigers or squirrels), although God won't classify those items together. He's also something of a fumbler when it comes to counting. (Again, this means God would think he's treating cardinalities of items as the same or different that God sees as different or the same.) Crusoe 3's senses and counting skills aren't *bad enough* to kill him—but bad enough that he often finds himself having a bad time of it. Finally, he *intends* his words not merely as his dispositions incline him to so apply those words, but as applying to certain collection of items in the world and not to others.

To make this work, we need to explain how Crusoe 3 is to understand *his* words "coconut," "tiger," "natural kind," and so on, so that *they* mean what he—according to this suggestion—intends them to mean. The first step, it will be said, is that Crusoe 3 can't understand the standards of his words to be as his own dispositions (or a subset of them) would incline him; he must instead understand the standards of his words to issue from the way things are in the world. But how is *he* supposed to manage this? All he has, after all, to make his kind terms match the kinds of the world correctly are his *actual* dispositions to apply these words—that's as far as his *natural* powers extend. How is the switch to be made from his actual dispositions doing this (and thus his dispositions being the *de facto* standards for what his words mean) to the things in the world being the standards instead? How, in other words, does the way the world is the standard for the correct application of his words manifest itself in his intention? He can't, for example, simply *think* or *say*: "the standard for the application of my word 'coconut' are the *real* coconuts." After all, "real," as in "real coconuts" or "real world" is one of *his* words. It seems that Crusoe 3, by somehow using any of his own words or any of his own concepts to refer to something beyond what he's disposed to apply them to, is attempting to "raise himself up into the air by pulling up on his own bootstraps."

Crusoe 3—like previous Crusoes—must take himself to be *relying* on his dispositions to group things as belonging to the "same kind," or as "similar," whenever he applies the word "coconut." What other option is there? And this means that he applies the word (or concept) "coconut" to whatever it is that strikes him as similar to *that*. Nevertheless, the suggestion is that he *understands* the meaning of his word "coconut" not as applying to whatever it is that *strikes him* as similar to coconuts, but as applying to whatever it is that *is* similar to coconuts. (But, given that this is to be *his* understanding, "is similar" must be *his* phrase as well.)

There are *two* problems we're facing here. One is how Crusoe 3 is supposed to *make* his words obey the world's standard. The second is how Crusoe 3 (using his words and his concepts) is even supposed to *think* or *understand* his words as the suggestion requires him to do.



Let's leave these puzzles aside for a moment. Notice that one necessary implication of this way of thinking that's being imputed to Crusoe 3 is that he will treat his specific characterizations of particular items as "coconuts" as *only defeasibly* applying (based as they can only be on his dispositions to apply "coconut"). He will have to regard himself as capable of *mistakes*. This notion of "mistake" is importantly different from the one that Crusoe 2 had, having decided to fault his tired disposition meanings in relation to his rested disposition meanings. Crusoe 2 describes his tired dispositions as mistaken by using his rested dispositions as the standard against which his tired dispositions are to be judged. Crusoe 2 recognizes mistakes as using the "wrong" words (or the "wrong" dispositions with the right words), where this means: using words or dispositions that are *less valuable* than ones he would have used if he were well-rested. Crusoe 3, however, is supposed to be intending something different: *the world* is to be the standard, not any subset of his dispositions. And this means that his mistake isn't in his choice of language, but is one about the *objects themselves*. He got something wrong about the *objects*, and how those objects should be grouped.

So we must return to our original questions. First, how do Crusoe 3's intentions enable his words "coconut," "similar," "natural kind," and so on, to bend to the "world's standard" (to, in other words, "match" the particular natural groupings in the world)? *Something* must be enabling this to happen, and so this way of posing the question is natural: What are the *facts* about the world (and Crusoe 3) that determine that his word (concept) "coconut" refers to this collection (all and only the *coconuts*), and not some other collection? What are the facts about the world (and Crusoe 3) that determine that his word (or concept of) "similarity" groups things in collections that correspond to kinds as they are in the world, and not in some other way? This question is made more vexing by the fact that if we follow the lead of Crusoe 3's dispositions, we will *not* get all and only the coconuts as what his word "coconut" refers to.

Second, how can Crusoe 3 even get his thinking (since it's in the medium of his own words and concepts) to be *about* "things as they are," independently of his own dispositions to use his words and concepts, so that he can even *entertain* the suggestion that the standards for his words are things "as they are," as opposed to things as his dispositions determine him to think of them?

Crusoe 3's supposed understanding of the meanings of his own words has created a (supposed) space between what his words "coconut," "natural kind," "real," etc., "actually" refer to, and his resources for attaching them to items in the world. And the questions focus on the supposed facts needed to determine—independently of Crusoe 3—what these words refer to. The problem is simply that Crusoe 3's dispositions *exhaust* his resources for determining what his words refer to; they exhaust what he can even think of his words as referring to. How can anything else help? (How can anything else change this?)

What a large number of philosophers in an influential tradition in metaphysics are tempted to say at this juncture is that *the world* does the job for Crusoe 3 because objects in the world *themselves* sort into natural kinds, and "coconut" as used by

Crusoe 3 refers to a natural kind.<sup>1</sup> One way it might strike philosophers that this could work is this: As long as Crusoe 3 can be sure that a *couple* of items are similar (belong to the same kind)—and surely *that's* not too much to demand<sup>2</sup>—and so long as items in the world behave nicely enough (sort into simple enough kinds, as I've stipulated to be the case with respect to the items in the worlds that all the Crusoes so far discussed live in), then Crusoe 3's word "coconut," for example, will refer (by his stipulation and by how things are in the world) to all and only the coconuts. Lewis (1983, 47) writes:

It takes two to make a reference, and we will not find the constraint if we look for it always on the wrong side of the relationship. Reference consists in part of what we do in language or thought when we refer, but in part it consists in eligibility of the referent. And this eligibility to be referred to is a matter of natural properties.

One (first) crude way to interpret the suggestion is this: The natural kinds in the world operate as "reference magnets": they semantically *pull* the references of Crusoe 3's words onto the natural kinds so that his word "coconut," for example, refers to all and only the items that belong to the same kind as his pair of sample coconuts do.

But the natural counterargument to the crude interpretation is this. Regardless of how neatly items in the world are organized (and to repeat, worlds don't get much neater than the ones I've described the Crusoes as living in), we still need to explain how *Crusoe 3* manages to get his word "coconut" to cover all and only the coconuts. It isn't somehow built into his words—it isn't built into *any* words—that (of course) they won't refer to irregular portions of reality. (These are *his* words—how was such a condition built into *his* words, and by whom?) The point can be put another way. The additional metaphysical facts about natural kinds won't help unless we are also told how Crusoe 3 is *sensitive* to these facts when he uses words like "coconut." The relevant sensitivity, of course, isn't a primitive "semantic sensitivity": semantic sensitivity has to be built out of our antecedent capacities to interact and determine aspects of the world—howsoever we do that.<sup>3</sup> Call this *the sensitivity demand*. Notice that I (pretty much) stipulated that the sensitivity demand is met in the cases of Crusoe 1 and Crusoe 2 by idealizing their senses. In the case of Crusoe 3, the sensitivity demand isn't met.

---

<sup>1</sup>Lewis (1983, 1984). As just noted, many philosophers have followed his lead. See, e.g., Sider (2009), especially 400–401, where suggestive language—like "reference magnets" and "ideal interpreters"—is invoked to rhetorically help the suggestion along. See my (2000), especially Part III, § 5, for an earlier version of the argument given in this section against this approach to reference. Lewis's approach involves grades of eligibility, and not a simple dichotomy between pure natural kinds (that are completely eligible for our terms to refer to) and unacceptably unnatural kinds (that are completely ineligible for our terms to refer to). I get to all this in due course in this chapter.

<sup>2</sup>Well, maybe it *is*. But I'm not going to press this objection now.

<sup>3</sup>A relevant analogy: There is a great deal more electromagnetic radiation out there than we can detect with our naked senses. Bees see ultraviolet light; and we don't. Despite this, were the ancient Greeks able to *refer to* ultraviolet light? No—and pretty much because they couldn't *see* it. We're able to do something the ancient Greeks weren't able to do. That's because we have instruments that can detect the radiation (and we have some theory about the stuff too). See Azzouni (2004a).

There is textual evidence that Lewis rejects the sensitivity demand. For he (1983, 54–55) writes,

We have no notion how to solve the problem of interpretation while regarding all properties as equally eligible to feature in content. For that would be to solve it without enough constraints. Only if we have an independent objective distinction among properties, and we impose the presumption in favour of eligible content *a priori* as a constitutive constraint, does the problem of interpretation have any solution at all. If so, then any correct solution must automatically respect the presumption. There's no contingent fact of psychology here to be believed, either on evidence or daringly.

And he (1983, 55) adds, some paragraphs later:

The reason natural properties feature in the contents of our attitudes is that naturalness is part of what it is to feature therein. It's not that we're built to take a special interest in natural properties, or that we confer naturalness on properties when we happen to take an interest in them.

As I've suggested, it looks like Lewis is rejecting the sensitivity demand; he's offering an *a priori* constitutive stipulation about reference (and content) instead of the revelation of real (for example, worldly structural) resources that supplement the individual's powers so that reference can be made determinate. The key phrases in what I quoted from Lewis above that support this interpretation are these: "There's no contingent fact of psychology here to be believed, either on evidence or daringly," and "It's not that we're built to take a special interest in natural properties, or that we confer naturalness on properties when we happen to take an interest in them." So, for Lewis, it suffices to state that the world's furniture sorts neatly into kinds, and that in turn suffices for our use of the world's-furniture-sorts-neatly-into-kinds fact as a constraint on admissible referents for our terms. No additional psychological facts about a purported sensitivity to that sorting are relevant.

But it's precisely this rejection of the sensitivity demand that makes Lewis's suggestion look like "an ad hoc response to a problem," or "the postulation of an irreducible 'semantic force'" (Sider 2011, 27). And this is why it's so natural for Putnam to abuse the suggestion the way he does. Lewis (1984, 67) defends himself from Putnam's attacks by saying:

I take it that Putnam classes the solution I advocate with solutions that rely on supernatural graspings or intuitions. He assimilates the view that 'the world ... sorts things into kinds' to the preposterous view that the world gives things their names (RT&H, p. 53)! Recently, he has called my talk of elite classes 'spooky' and 'medieval-sounding'.[...] Well, sticks and stones may break my bones .... Anyway, what's wrong with sounding medieval? If the medievals recognised objective joints in the world—as I take it they did, realists and nominalists alike—more power to them. But I don't suppose that egalitarianism of classifications is an especially medieval notion—rather, egalitarianism is a peculiarity of our own century.

This response isn't pertinent to the objection—that no *sensitivity* to the world's natural-kind divisions is on offer. Remarks that Lewis has made that do seem to be pertinent to the sensitivity objection lie elsewhere, as I'll indicate when I discuss the other two interpretations of his views.

*Interim Conclusion* It's true: there is no way to make sense of the idea that Crusoe 3 can *make* any word—no matter how fundamental—pick out a something, or a kind of somethings, in the world by his sheerly intending it to do so, and by (somehow) having the world do the rest of the work needed. There is no way to even make sense of the idea that Crusoe 3 can *think* of doing this. It should also be clear that none of these conclusions depend on Crusoe 3's isolation. A *community* of people—all busily pulling up on their own bootstraps—are not more likely to rise into the air, than an isolated Crusoe 3 so pulling is.<sup>4</sup> Finally, it won't help to introduce additional *ad hoc* metaphysical devices—such as “properties,” or Fregean “concepts,” or other sorts of *abstracta*—that Crusoe 3 is to supposedly “grasp” and intentionally associate with his words, and that do the needed referential work for him. Why on earth should it be that he “grasps” just the kinds of prophylactic devices needed, and not other ones that fail to “match” the world in the contours they impose on Crusoe 3's words?

Notice that this objection *isn't* either of the two objections I quoted from Sider (ad hocness or the postulation of an irreducible semantic force)—or, rather, those objections follow as corollaries from the more fundamental one I'm pressing. This is that what Lewis has offered is a non sequitur—so far as the suggestion has been described, anyway. Make the world as rich in structure as you want; that's sheer metaphysics and not at issue. For it's a presumption of the very inquiry that leads to the rule-following paradox that we're in search of an *explanation* for why someone is counting rather than quounting, or referring to coconuts rather than some other unpleasantly unruly collection of items. This means that the sensitivity demand must be met: the postulated structure has to be one *we* can pick up on (that we have antennas for, as it were). If Lewis is an opponent of the sensitivity demand, he should be explicit about why—he should motivate rejecting it. I consider an interpretation of Lewis in the next section that takes him to be doing just that.

### 4.3 A Top-Down Approach to Reference-Magnetism: Williams's Interpretationalism

Sider (2011) has belatedly become scornful of the interpretation of reference magnetism that I've given in Sect. 4.2 (and that he offers in his (2009)). This is even though he still describes the suggestion that “predicates must stand for natural properties and relations in a correct interpretation” (p. 17) as “externalist” in the sense that, “reference is not determined merely by use ... [a] second factor [is that natural] properties and relations are ‘intrinsically eligible meanings’; they are ‘reference magnets’.” That is to say, Crusoe 3's word “coconut” picks out all and only coconuts because the set of all and only the coconuts *is* an intrinsically eligible meaning, and the alternatives aren't. Sider, as I mentioned in the last section, notes that some think

---

<sup>4</sup>To extend this metaphor: and if they pull on the bootstraps of *one another*, some people can be pulled into the air—but only in relation to others who firmly remain on the ground.

this Lewisian suggestion is ad hoc, and others think of it as the “postulation of an irreducible ‘semantic force’.” “Neither charge is founded,” Sider (2011, 27) writes, because we can approach the notion of reference magnetism the way Williams (2007, section 2) does, by deriving it “from a well-motivated and more general doctrine about theoretical virtue.”<sup>5</sup>

This is a good strategy—but only if it can be made to fly. The overall contours of the strategy are these: Good scientific theories have an optimal combination of several virtues, simplicity, (deductive) strength, and fit (with the data) among them. In turn, simplicity requires the primitive terms of a *scientific theory* to hold of (relatively natural) properties and relations. “Relatively natural” because, although (on this second view being attributed to Lewis) fundamental physical theories must have terms that pick out fundamental (perfectly natural) properties and relations, special sciences will have terms that only refer to relatively natural properties and relations. Since semantics is a special science, it inherits the simplicity requirement (that holds of the sciences generally); and so its terms—e.g., “refers”—must also hold of relatively natural predicates and relations. This, finally, will determine that *our* semantic theory of Crusoe 3 requires that his word “coconut” refers to all and only the coconuts, and his word “counting” refers to the relatively natural relation of counting (as opposed to the relatively unnatural relation of quounting).

Notice that this view doesn’t attribute any supernatural sensitivities to Crusoe 3 (or to us). It isn’t that natural kinds and relations themselves *play* a role in fixing the references of our terms. (So the nomenclature “reference magnetism,” as well as much of what I quoted from Lewis above, is misleading—if this second way of interpreting “reference magnetism” is to be attributed to him.) Rather, there is a constraint on how the special term “refers” works, because it’s a term of the special science of semantics. It must pick out a relatively natural kind, *relative to other scientific virtues*, e.g., fitting the data. One appealing point about this interpretation of Lewis is that the ominous language that he uses—“a priori,” “presupposes,” and “constitutes”—is also inappropriate. Instead, the constraint in question is simply an empirical one; it issues directly from ordinary scientific practice.

The drawback of the strategy, of course, is that it turns on getting *actual* ordinary scientific practice *right*. (It thus provides a “hostage to fortune.”) If actual ordinary scientific practice isn’t as Williams (and Lewis, on this view) depicts it, then the strategy fails on those grounds alone.

As my preliminary discussion makes clear, Williams engages in two projects in the relevant article: motivating his particular form of interpretationalism—but also argu-

---

<sup>5</sup>There is irony in Sider’s invocation of Williams (2007). For Williams makes clear in that article that he thinks the approach that Sider likes ultimately fails—it faces a fatal-looking indeterminacy argument. I won’t pause to evaluate Williams’s clever argument against his interpretation of Lewis’s version of reference magnetism (since there are more basic problems with it). I only want to officially announce that I don’t see why Sider’s invocation of rich metaphysical structure provides any resources for meeting Williams’s objection to this version of reference magnetism. (Sider (2011), notably, doesn’t even indicate that Williams thinks the approach Sider endorses won’t work; consequently Sider doesn’t tell us how he would meet the objection if he did acknowledge it.)

ing on the basis of text that the view is Lewis's. I'll largely set the second issue aside and focus only on whether Williams's interpretationalism can work.<sup>6</sup> I'll argue it can't.

It's an analysis of the theoretical virtue of simplicity that drives the philosopher of science (according to Williams) to the position that a graded division between natural and non-natural kinds is needed. To see this, consider two theories of micro-physics, *T* and *T'*, where *T'* "is just like *T* except for the addition of a 'redundant' natural law: one that generates no new predictions about particular matters of fact" (Williams (2007, 371)). For concreteness, imagine that the extra theoretical tissue in *T'* "governs the behavior of particles under nomologically impossible circumstances: what an atom would do if it traveled faster than light ..." (Williams (2007, 371)).

There are different views of what makes *laws* true (or true generalizations *laws*). According to Humeanism, the laws of a scientific theory being true involve nothing additional to what's involved in making any other statement true: a law is true if and only if it fits "with matters of particular fact" (Williams (2007, 372)).<sup>7</sup>

But if all that's involved in theory selection are matters of fact that theories are to fit with, then *T* and *T'* look on a par. As Williams (2007, 372) writes, "Call this the argument from additional junk: it threatens to show that Humeanism about laws of nature will make it indeterminate whether or not the redundant law included in *T'* holds." Notice the simple logical point behind this concern. Vacuous generalizations like,  $(x)(Gx \rightarrow Hx)$ , where *G* is uninstantiated, are true. But are they therefore *laws*? If the Humean places no additional constraints on lawlikeness other than the usual truth-functional machinery, then the answer has to be "yes."

So Williams concludes (on behalf of Lewis, and on behalf of Humeans generally) that more has to be involved. There are, of course, other theory-relative virtues to reach for: simplicity for example (how economical and parsimonious the theory is), and (deductive) strength. The best scientific theory "is one that has the optimal combination of simplicity, strength, and fit" (Williams (2007, 372)).

Lewis analyzes simplicity in terms of syntactic complexity (after regimenting a scientific theory formally). According to Williams (2007, 373), "count an axiomatized theory as simpler than another if it has fewer, and syntactically less complex,

---

<sup>6</sup>I'll point out in this footnote, however, that I doubt Williams is right about Lewis. Williams structures Lewis's view so that what Lewis thinks about science (and laws) applies directly to semantics because semantics is a special science; and in this way Lewis derives his purported interpretationalism from his broader views about science. But there is no textual evidence that Lewis derivationally structured his views this way. (And Williams gives none. On the contrary, actually—see, in particular, Williams (2007, 371), footnote 21. No new textual evidence is given by Williams (2015), incidentally.) By contrast, I see Lewis's use of syntactic simplicity, for example, as a general-purpose tool he reached for whenever he needed it. Nothing significant turns on figuring out Lewis's real views about these matters. I treat Williams's suggestion the way I treat Kripke's interpretation of Wittgenstein: as an interesting and important suggestion to be evaluated on its own merits, regardless of whether it holds of the original philosopher it's attributed to.

<sup>7</sup>Humeanism is the background position assumed here: Lewis holds the view, Williams is working within its parameters and, for the record, I'm a Humean too—subject, however, to the significant caveat that I think sentences can be true without term-world correspondences. This caveat matters in Chap. 6 but that issue can be set aside for the purposes of the current discussion.



axioms.”<sup>8</sup> Unfortunately, there are tricks available to transform a theory, no matter how syntactically complex, into an equivalent one that’s maximally syntactically simple.<sup>9</sup> The way around this is to require a gradation of natural properties. A second crucial move occurs right here. Williams (2007, 373) writes:

In order for [the introduction of natural vs. unnatural properties] to contribute to an *objective* (partial) analysis of simplicity in terms of syntactic complexity, we need to postulate an objective distinction between the natural/non-natural—between elite properties and abundant rubbish.

Up until this point, the discussion has involved *purely theoretical* virtues—which is as it should be. The Humean wants distinctions among laws to be *objective*—but fully derived from properties of theories. Now (all of a sudden, as it were), we are to introduce a *metaphysical distinction* in the world—apparently motivated by the desire to avoid “subjectivity.” We’re forced to this because syntactic construals of simplicity fail; and we don’t want to acquiesce in the idea that simplicity is a partially subjective notion, based on a projection of what seems simple “to us” (Williams, (2007, 372)). But this move deserts the Humeanism that it supposedly supports. We are now postulating genuine metaphysical structure that plays a crucial role in distinguishing *T* from *T'*. What’s wrong with the Humean, instead, treating *T* and *T'* as both (accurately) describing the relevant regularities (which they do), and *T'* being ruled out simply because it distractingly involves unreal cases (a “subjective” matter, I guess—but one that’s crucially connected to the degree of user-friendliness a theory exhibits)? In the same way, the Humean can simply accept that vacuous generalizations are laws—but pretty useless ones. Why is the postulation of a rich metaphysics *better* for a *Humean* than admitting the existence of useless vacuous laws? Why is the admission that numerous laws are useless so much

---

<sup>8</sup>Williams (2007, 375), footnote 27 “sets aside” issues of measuring syntactic complexity. But all attempts to manage this are known to have failed for so many reasons, including the various major technical obstacles they face—and that Williams alludes to. I’m also thinking, particularly, of the fact that formal simplicity has nothing much to do with ordinary scientific views about how one theory is simpler than another. I pass over all this tangled philosophy of science because conceding the point to Williams (and Lewis, on this interpretation) won’t help their case. I’ll note that Williams (2015, 373) expresses firmer opposition to the idea; he writes: “I don’t really have a clue what a “definition” of the ordinary subject matter of thought and talk—shoes and string and sealing wax—would be, if the definiens is to be drawn from microphysics. I suspect you don’t either. There’s no reason to believe that such definitions are even possible.” The problem, of course, generalizes to the relationship between the special sciences and microphysics—and even between microphysics and other branches of physics! (I discuss this point further in Sect. 4.4). Williams (2015) explores other options—but comes up short.

<sup>9</sup>Williams’s discussion of this point is tortured. He admits that the trick he describes isn’t conclusive because the simple theory the trick yields has no deductive powers, and surely deductive power is a theoretical virtue as well. Nevertheless, ruling out syntactically-simple axiomatizations (of theories) due to this kind of trick is the *crucial* justification—on his interpretation of Lewis—for supplementing the notion of simplicity with a distinction between “perfectly natural,” or “fundamental” properties, and “the rest.” But if this is right (if this is all Williams has to justify his wedding simplicity to metaphysical naturalness), Williams has in effect admitted (explicitly in footnote 23: “the case is not decisive”) that his justification for the natural/non-natural distinction fails.

worse for the Humean than positing a rich chunk of metaphysics, and rooting the distinction between laws and mere generalizations in *that*?<sup>10</sup>

Bypassing the objections of the last two paragraphs, how does the natural/non-natural metaphysical distinction impact the special science of semantics—how in particular does it get us reference magnetism? The idea is this. Once one has a measurement of the simplicity of axiomatizations (in terms of syntactic complexity as well as in terms of an eligibility—naturalness—measure of what the primitive terms refer to), the resulting simplicity virtue can take its place *among the other* scientific virtues when doing science—specifically semantics.

So consider the question whether we're counting or quounting—the question, in other words, whether a semantic theory that takes our term “count” to refer to counting is simpler than one which takes our term “count” to refer to quounting. We're supposed to draw the conclusion that positing a referential relation between “count” and counting is simpler than positing one between “count” and quounting—and part of the brief for this claim is that counting is a more natural (mathematical) relation than quounting.

Even if we grant the greater simplicity of an axiomatization of semantics that treats “count” as referring to counting, we still don't get what we want. The reason is that when Lewis speaks of the reference to (relatively) natural properties as an *a priori* constitutive constraint on content, other scientific virtues (like “fitting the facts”) look like they're *ruled out of court*. But even after granting all the moves Williams makes—many of which I've challenged—the result is still only the weaker one that natural properties function within the context of a simplicity calculation as a scientific virtue *along with other scientific virtues* to be used to judge a theory optimally good. That's a (dialectically) far weaker position than what Lewis seems to be (rhetorically) striving for.

In particular, the threatening rule-following paradox is *back*, but now in the guise of a constraint on fitting the facts, and in this way all the psychological facts that Lewis explicitly rules out (in what I quoted from him above), also return as relevant. For the psychological facts about us that show we have no resources to distinguish adding from quadding and counting from quounting are ones—it seems—relevant to the factual data that a semantic theory should take account of. We *didn't* reject general relativity because it's so much less simple than Newtonian physics—although it is. The history of physics, in fact, is the history of the substitution of progressively more complex theories for less complex ones because the simpler theories don't “fit the facts.” Are semanticists instead allowed to be *that* lazy? Is that the prerogative of practitioners in special sciences? (I *hope* not.)

*Interim Conclusion* Williams's reconstrual of Lewis' approach fails for an interesting reason. It fails because if naturalness is construed as a component of simplicity, its role in semantics has already been taken account of by the meaning sceptic who

<sup>10</sup> Here's a reason to think positing a rich chunk of metaphysics is better: Laws are to operate counterfactually, and laws with empty antecedents don't have empty antecedents in certain possible worlds. But the best *Humean* solution to *that* problem seems to be to stipulate conditions on when possible worlds are relevant to the counterfactuals rather than messing with the *actual* metaphysics to exclude laws with excess (but vacuous) conditions.



is posing the rule-following paradox. For it's an important assumption of (naturalized!) semantics that if we refer to something, it's because we have resources for doing that. So the "fitting the facts" virtue trumps simplicity in this case—as it so often does in the sciences.<sup>11</sup>

It's an important methodological point that scientists often use a theory they take to be false because it's simpler than the true theory.<sup>12</sup> But this falls under the general rubric of "idealization." That's not the suggestion Williams and other proponents of reference magnetism are making. They want to tame the rule-following paradox, not concede its force by treating the introduction of graded natural properties as involving a false but usefully simple semantics.

#### 4.4 Moorean Considerations Plus the Only Game in Town Argument

Lewis's position is open to another interpretation—a third one—that's far more intransigent than Williams's interpretation. This is a combination of an acknowledgement of Moorean facts along with a version of the only game in town argument.<sup>13</sup> I'll begin with a very brief indication of how Lewis's discussion motivates this interpretation of his position.

First, Lewis (1983, 46) treats Putnam's radical indeterminacy argument as refuting "the supposition that there are no further constraints on reference" other than what the individual thinks or says. Presumably the same lesson can be drawn from the rule-following paradox. Lewis adds (1983, 47), "Our language does have a fairly determinate interpretation (a Moorean fact!) so there must be some constraint not created *ex nihilo* by our stipulation."

The following view can be attributed to Lewis on the basis of these quotations; a number of philosophers have taken him to have this view; and many have adopted it themselves (or opposed it).<sup>14</sup> First, standard (successful) semantics requires deter-

<sup>11</sup> This, incidentally, gives a second reason to think that Williams's reconstrual of Lewis's approach isn't Lewis's original one. Lewis, as I indicated in Sect. 4.2, seems to reject the sensitivity requirement; but Williams's reconstrual requires taking account of it.

<sup>12</sup> See Azzouni (2000), especially Part I, § 2, for discussion of this, and for citation of other philosophy of science literature that makes the same point. But of course the continued and widespread use of various aspects of Newtonian physics is an example pretty much *everyone* knows about.

<sup>13</sup> Fodor (1975) famously—among philosophers, psychologists and linguists, anyway—uses the phrase "the only game in town" and runs versions of this argument on behalf of various theses in philosophy of mind—especially RTM (the representational theory of mind). One of his cutest allusions to the argument is his (1975, 27) quotation from Lyndon Johnson: "I'm the only President you've got." The only game in town argument, of course, is a version of inference to best explanation—but in the special case where it's argued only one explanation is available.

<sup>14</sup> I think this third position—one that should be clearly distinguished from the other two—is really Lewis's. Williams's interpretation of Lewis is ruled out for the reasons I've already given. The crude view of Sect. 4.2 is ruled out because it's, well, crude.

minate interpretations for our words—and not the referential indeterminacy that rule-following threatens us with. Second, because of this, standard (successful) semantics requires a graded distinction between natural and unnatural properties. Third, this is the only game in town. In this section, I dispute the first and third assumptions.

Let's start with the natural question: what exactly *is* the Moorean fact about the determinate interpretation of our language? Well, ironically, it's just not obvious—so we have to go through the possibilities. There are, first, the humdrum facts about the reference practices of a population. These facts actually seem more like data *for* the special science of semantics than serious theoretical doctrine that belongs *to* it. I mean things like this. When Crusoe 3 calls something a coconut, it often is a coconut. Not invariably, of course, because he often picks up things that he calls “coconuts” but later denies are coconuts (and, he might even reverse himself again because mistakes in this arena of behavior are often like this). These are Moorean facts, if anything is.<sup>15</sup> There is also the wealth of practices surrounding counting: what we say as we count, and notably what we say we *know* (and don't know), and again, how we correct ourselves (and others) when we make what we call “mistakes.”

This, let's say, is a gigantic body of verbal and nonverbal Moorean facts. But (unfortunately for anyone who wants to run this Lewisian line), it's hard to see how any of this can be the stuff Lewis is talking about—after all, all of this is precisely what Kripke's meaning sceptic pores over so carefully, and uses as the basis for his rule-following paradox. So these particular Moorean facts are common coin between these opponents. They don't especially help the Lewisian.

So let's try again. Could it be that the *generalizations* (that one finds in standard classical semantics) are the Moorean facts that Lewis is alluding to when he describes our language as having “a fairly determinate interpretation”? (Isn't it times like this that you become particularly irate at how, um, *vague* the pronouncements of philosophers can sometimes be?) What do these purported generalizations look like?

Well, again, we have to guess, since Lewis hasn't told us. One straightforward set of candidate-examples (they're a not generalization, exactly, but collectively close to one) are the crucial Tarskian biconditionals—sentences of the form:

“Snow is white,” is true if and only if snow is white.

That is, all sentences that result from replacing, respectively, “snow is white” and its standard quotation-name in the above sentence with any other sentence and *its* standard quotation-name.

But these won't help Lewis at all. The problem is that the constraints offered by even infinitely many Tarski biconditionals are too crude to fix the references of the

<sup>15</sup> Well, qualifications are called for because these are facts about *usage*, and they turn out to be far more intricate than Lewis realized. They're also generally invisible to natural-language users (which makes them bad candidates for *Moorean* facts), and this invisibility point is one Chomsky makes regularly. Finally, they also strongly resist treatment via the standard semantic theories that Lewis likes. For details about some of this, see my discussion of Chomsky and Pietroski later in this section. Also see Azzouni (2013), where all this is discussed at more length.

terms that appear in them. These biconditionals (being biconditionals) merely coordinate the truth values of two sets of sentences—but that isn't enough to fix the references of specific terms. This problem, of course, generalizes if the semantic theory we have in mind is one of a family of important semantic theories influenced by Tarski's work: truth-conditional semantic theories. This is because the primary constraints on the references of the terms of such theories directly derives from the Tarskian biconditionals, and those—it's already clear—are insufficient for Lewis's purposes. The other constraints imposed by such theories are compositional ones—but these don't offer additional constraints on the references of terms—not ones sufficient to provide determinate references, anyway.<sup>16</sup>

The situation is worse than I've so far indicated it to be for this brand of Lewisianism. It can be shown that even those truth-conditional semantic theories that successfully capture our use of context-sensitive expressions ("this," for example, as in "This lamp isn't lit.") don't determine reference in the way the proponents of this reference-magnetism approach require.<sup>17</sup> This makes the situation worse, of course, because these context-sensitively phrased Moorean facts about how we refer seem to be ones that aren't simply "more language" that's open to reinterpretation but something stronger: they look like real contact with the world. If Moorean facts about real contact with the world won't help the Lewisian, whatever could?

Other sorts of semantic theories, ones that break with the truth-conditional-semantics tradition, also fail to provide the Moorean facts that Lewis needs. Consider, for example, the "structural semantic theories" of Katz (1972), Katz and Fodor (1963), and Jackendoff (1972). Such theories are designed to provide mappings of each natural-language expression to those of a representational language, either a language of thought (as in Fodor 1975), or an abstract language (as in Katz 1981). Leave aside that Lewis certainly *doesn't* have these kinds of semantic theories in mind because he (1970) argues that such mappings provide no interpretation of a natural language: these mappings are from one system of notation to another, Lewis complains, but that tells us nothing about what the terms of either system are about.<sup>18</sup> The important point is that it's clear (and Lewis thinks this as well) that such theories can't supply determinate worldly references for the terms of natural languages. That's not what they're in the business of doing.

It might be argued that I haven't focused on what Lewis means when he writes, "Our language does have a fairly determinate interpretation (a Moorean fact!)." But what did I miss? I've covered the humdrum facts about our referential practices; I've covered the semantic theories that characterize those practices; I've even extended the discussion to semantic theories of languages, where they connect behavioristi-

<sup>16</sup> If they did, then the theorems that Putnam's indeterminacy results rely on would fail—permutation results and the Löweinheim-Skolem theorems. See Azzouni (2003) for the technical details.

<sup>17</sup> See Azzouni 2010b, chapter 5 for details on this. I also revisit the topic in Chap. 6 of this book.

<sup>18</sup> Whether this observation really infirms such semantic theories can be left aside. See Azzouni (2010b), section 5.3, for an argument that such semantic theories are perfectly good (semantic) theories despite what Lewis (1970) and other critics, e.g., Higginbotham (1990) and Lepore (1983), claim.

cally to gestures we make towards the world itself, as when we point at a lamp and say it's not lit. It's hard not to draw the conclusion that Lewis's vague allusion to "determinate reference" and "Moorean facts" begs the question against his opponents.

It's *almost* not worth making the following point. I've been undercutting the suggestion that Moorean facts can be used by Lewis against his meaning-sceptic opponents. It follows that it won't help to pile these Moorean facts together into a gigantic theory. Call this the *many bootstraps principle* (I invoked it in Sect. 4.2): if one item pulling on its bootstraps can't rise into the air, a few thousand of them doing something similar won't induce a miracle either.

Two last points need to be made to round out this section. The first is that I've sketched my argument against Lewis's use of Moorean facts at a certain level of abstraction. More evidence against this maneuver is coming in the following two chapters. For I'll show there, using later versions of Crusoe (psychologically realistic ones or, at least, Crusoes who seem much closer to us), that a successful language practice coupled with attributions of truth and falsity to the sentences made in that language further coupled with a standard truth-conditional semantics for the language, is completely compatible with "indeterminate reference." More specifically, a concession to Kripke's meaning sceptic about a language has no effect on truth attribution or the attribution of a standard semantics to that language. These will be *examples* of why Lewis's invocation of Moorean facts doesn't help him.

A second point should be made about the status of semantics as a science. This will take a little while to develop. I alluded several paragraphs back to a non-truth-conditional tradition in semantics in which, consequently, reference isn't a central notion. This is the tip of an iceberg of a much more dramatic challenge to Lewis's approach to the rule-following paradox. What's at issue is whether, if and when semantics emerges as a science, a determinate reference relation will play a role in it; for it's only a determinate reference relation that Lewis can use against the meaning sceptic.<sup>19</sup> Notice what's needed: The purported science of semantics has to be one in which determinate reference plays a central "explanatory" role—where it operates as a linchpin "theoretical" notion.<sup>20</sup> Sider (2011) writes that "some philosophers reject reference-based explanation." He adds that "some claim that conventional meaning encodes far fewer referential properties than mainstream semantics thinks, but even they usually grant *some* explanatory role to reference—or at least to reference-by-a-speaker," and then he cites Chomsky (2000) and Pietroski (2003) as illustrations.

This is a serious misreading of Chomsky and Pietroski, and a serious misreading of how fundamentally opposed they are to anything in the ballpark of "reference-based explanation." First of all, "reference-based explanation" is an allusion to truth-conditional referential semantics in Tarski's style (as the phrase "mainstream semantics" indicates). But Chomsky and Pietroski (as well as a large number of lin-

<sup>19</sup> This applies not only to the Moorean Lewis that I'm currently discussing, but also to the more moderate interpretational Lewis of Sect. 4.3.

<sup>20</sup> Davidson presumes this about "reference." See, e.g., Davidson (1973, 74).

guists) definitively reject this approach to semantics. The question, for both Chomsky and Pietroski, is whether the terms of natural language are empirically amenable to anything like the notion of reference, as used in this “mainstream” tradition. Chomsky adamantly rejects the idea in the book that Sider refers to. He shows that natural-language names and kind terms are nothing like the terms amenable to a Tarski-style reference relation. Pietroski (2003) makes the same point with related examples.<sup>21</sup>

Pietroski (2003, 222) makes the often repeated observation that people refer and not words. This slogan, if not read carefully in context, may mislead because it can give the impression that the natural-language terms resistant to standard-semantics treatments can be handled by substituting term-reference-as-used-by-a-speaker, or term-reference-in-a-context-as-used-by-a-speaker, for term-reference. Indeed, Sider’s suggestion that “reference-by-a-speaker” is taken seriously by Chomsky and Pietroski misses the force of their objections because they make it clear no simple expedient of this sort handles the examples.<sup>22</sup> The point being made by Chomsky and Pietroski is that whatever emerges as a *science* of semantics is unlikely to involve anything like determinate reference of the sort required by Lewis’ response to the rule-following paradox. This is because any determinate referential relation, like the truth and falsity of the sentences we utter, is (as Pietroski (2003, 218) insightfully puts it) an “intractable interaction” effect. It will not, that is, play an explanatory role in semantics; it will, instead, be rather like the concept of “friction” in physics—something that’s *to be explained* as the resultant of *many* effects, some semantic, and many not. It will not itself be a useful theoretical notion.<sup>23</sup>

The case that Chomsky and Pietroski make against reference-based explanation turns on ways that names and kind expressions can be empirically shown to operate in natural languages—ways that are strongly counter to what reference-based semantics requires. Apart from these problems, there is a second way that kind terms—in particular—seem to violate naturalness constraints. Kind terms in the

<sup>21</sup> See Chomsky (2000, 37, 180–1, 35–36), Pietroski (2003, 226–233).

<sup>22</sup> Consider Chomsky’s (2000, 37), “London is so unhappy, ugly, and polluted that it should be destroyed and rebuilt 100 miles away.” Notice that “London” is functioning in a way that resists standard Tarskian reference-treatment in a *single sentence*. This is a phenomenon rampant in natural language: a single use of a name—“London”—taking in a single sentence predicates that presuppose disparate referential targets. Also see Pietroski (2003), on various attempts to modify truth-conditional-semantic approaches to handle these examples *systematically*.

<sup>23</sup> Chomsky often directly attacks the standard notion of reference—describing it as useless for linguists, e.g., in Chomsky (2000, 40). For discussion of this aspect of Chomsky’s thinking, see Azzouni (2013). Note the contrast with the opening paragraph of Williams (2007, 361); he writes, “The subquestion I focus upon here concerns the semantic properties of language: in virtue of what does a name such as ‘London’ refer to something or a predicate such as ‘is large’ apply to some object?” Notice also that there are two issues with the notion of reference. One is how it’s to be explained (or not explained) in the context of semantics; a related—but importantly different—idea is the theoretical/explanatory role that “reference” plays in the science of semantics. When Sider talks about “explanatory role” he’s focused on the second idea; the opening paragraph of Williams (2007) is officially focused on the first. In Tarskian approaches, they dovetail together. In the approach to semantics urged by Chomsky and Pietroski they are both largely sidelined for the same reason: theoretical intractability.

special sciences and in ordinary life seem to pick out properties that are relatively unnatural—at least in relation to the terminology of physics. Lewis is explicit about the issue, and he introduces graded naturalness to handle it, where the gradations result from the syntactic complexity of the definitions of the less natural properties in terms of the more natural properties.

Lewis (1984, 66), when speaking of the relationship of the most elite properties (those of physics) to those less elite, e.g., “sticks and stones, cats, books, stars,” writes:

The less elite are so because they are connected to the most elite by chains of definability. Long chains, by the time we reach the moderately elite classes of cats and pencils and puddles; but the chains required to reach the utterly ineligible would be far longer.

There is a lot of independent literature on the possibility of such (chains of) definitions; it arises, naturally enough, with respect to the issue of the reduction of the special sciences to underlying sciences.<sup>24</sup> The problem, sometimes described as the multiple-realizability problem, is that special-science terms seem to be unnatural disjunctions of underlying terminology. Lewis, apparently responding to this idea, as I indicated by the quotation from him above (and Williams (2007) following him), offers a definitional syntactic solution to the issue.

This is a little surprising because the (philosophy of science) literature recognizes that definitions are simply not an option here. (This was realized and publicized as early as the 1960s.) That knowledge, among other things, has driven research on supervenience, for example, as a tool to anchor the relationship of special-science terminology *semantically* to underlying terminology without providing actual definitions—since the latter aren’t forthcoming. Whether, in fact, the Lewis approach to graded naturalness can be modified by dropping the syntactic criterion for naturalness and substituting one or another semantic criterion isn’t something I’ll pursue now. This is because there is still a worse problem.

The worse problem is that it’s clear, again by means of counterexamples, that in fact the terminology of ordinary life and the sciences aren’t amenable to naturalness considerations because, instead, term-usages in these areas are clearly sensitive to surface phenomena such as appearance-properties and (more ominously) contingent historical facts about the order of discoveries. Sider (2011, 32) acknowledges simple examples like “amulet”—as used by a linguistic community to refer to all items of gold and one particular silver item. And this kind of example looks completely amenable to a graded-naturalism approach. But there are numerous examples from all the sciences, both real and thought-experimental, that show that “joint-carving” issues aren’t particularly significant for natural-kind terms.<sup>25</sup> The problem, as always, is the massive contingency of introduced scientific terminology. If, as LaPorte (1996, 117–20) points out compellingly, we had discovered a planet like

<sup>24</sup> See Kim (1993), Putnam (1967), Fodor (1974). See Azzouni (2010b), section 4.2, for discussion and the citation of additional relevant literature.

<sup>25</sup> See LaPorte (1996), Dupré (1981), Azzouni (2000), and Wilson (1982), for examples. Sundell (2012) makes a similar point.

Earth but with heavy water wherever we have water, and if this had happened before we discovered isotopes, the radical properties of heavy water (it kills fish, crabs, aquatic plants, it freezes at a different temperature, it can be separated from ordinary water, and so on) would have precluded an identification of it with water—even *after the discovery of isotopes*. The semantics of kind terms—scientific and otherwise—is likely to be very complex; it’s not very likely to involve naturalness at all.<sup>26</sup>

## 4.5 Concluding Remarks

I’ve done the following in this chapter. I’ve analyzed three possible approaches to “reference magnetism” (all derived from David Lewis’s work, but all—in various combined forms—very popular both with proponents and opponents). One version really does take natural kinds to exert some kind of semantic force on the referential properties of our terms, a second version treats graded naturalness as part of the bundle of virtues that all scientific theories (including semantics) should exhibit, and a third attempts to show that no other options (for semantics) are available. The counterarguments to these positions should now be clear. The first interpretation really does deserve the scorn Putnam (and others) have heaped on it, the second fails because (i) it’s *not* in accord with scientific practice and (ii) because, in any case, it does *not* yield the needed result that the naturalistic concerns about referential resources raised by the rule-following sceptic are trumped by simplicity considerations. The third interpretation fails simply because there are several competing games in town (all of which honor the Moorean facts of referential usage and even Moorean semantic laws—if there are any), but that reject the centrality of determinate reference that’s a core of “mainstream” semantics.

I should note this. Dorr and Hawthorne (2013, 3) describe an on-going debate between “naturalist enthusiasts” and “naturalist sceptics”—where this difference is glossed in terms of differing “attitudes” between those who “‘reject’ the distinction between natural and unnatural properties, while the enthusiasts are said to ‘accept’ or ‘countenance’ it.” They go on to deplore the fact that this disagreement is structured around “autobiographical claims.” Whether or not this accurately describes any debate in current metaphysics—I rather doubt it does—notice that the objections of this chapter have pretty much nothing to do with questions about how rich (or not) the metaphysics of the world is. The problem focuses, rather, on (i) the question of the resources available to agents to take advantage of a posited metaphysics to fix what they refer to, and (ii) the question whether that posited metaphysics should be built into theories about those agents.

In the next chapter, I resume my analysis of the rule-following paradox by means of the successive construction of ever-more complex Crusoes.

---

<sup>26</sup> For further discussion, see Azzouni (2010c), section V.



## Chapter 5

# How Positive Success Curves Enable Private Rule Following

**Abstract** Two more isolated rule-following Robinson Crusoes are discussed. Crusoe 4 still recognizes himself to be speaking a disposition-meaning language because he has introspective access to the dispositions that generate his meaning-urges. But because those dispositions change relatively continuously, he cannot think of himself as speaking distinct languages. Instead, he thinks of his language as one in which he has better and worse dispositions at different times. Crusoe 5, however, has no or little introspective access to the dispositions that generate his meaning-urges. I show that, therefore, he must be like us: he must think of the terms of his language as obeying standards due to the way objects are in the world instead of as due to his dispositions. In this chapter, I explore what this means.

### 5.1 A Robinson Crusoe with Continuously Varying Dispositions that He Has Conscious Access to

Let's now consider a Crusoe 4. He lives in the same kind of artificial world that our previous Crusoes of Chaps. 2 and 3 live in. Like those Crusoes, he recognizes himself to speak a disposition-meaning language where his current meaning-urges are what his words currently refer to. Like the previous Crusoes, Crusoe 4's memory (by God's standards) is very good; and so he can remember both the specific items that he applied his words to in the past, and the states he was in when this happened. Unlike Crusoe 2, who has only two static sets of urges, Crusoe 4's dispositions steadily evolve over time. They, further, exhibit a positive "learning curve." That is to say, the meaning-urges that he uses to apply his words steadily improve over time—according to God's standard of how well they match the natural kinds of the world. From God's point of view Crusoe 4 gradually, but continually, gets better at identifying, distinguishing, and making similarity judgments about the objects in his world.

This is how *God* sees it. What *Crusoe 4* notices, instead, isn't a positive *learning* curve. Rather, over time, he experiences new internal states—new meaning-urge states—in which his urges to apply words enable him equal or more success in getting around his world than did any of his previous meaning-urges, in previous internal states. That is, Crusoe 2, recall, had two internal states he could recognize (according to God's standards): being tired and being rested. Crusoe 4, unlike Crusoe

2 and unlike us, recognizes that he has many internal states that bear on his meaning-urges, he can distinguish them from one another, and he recognizes that he's often in new internal states that he hasn't experienced before. Because the new internal states seem to always accord him more success than previous internal states, he's adopted a systematic policy of always avoiding (as much as possible) doing anything—hunting down coconuts, for example—when he's not in a new internal state.

Crusoe 4 doesn't always have *choices* about this, of course. (He has to run away from tigers if they're chasing him—no matter what internal state he's currently in.) But he has found that he can induce superior internal states if he behaves in certain ways. For example, he can avoid drinking, he can try to have experiences in good light or when things around him are relatively quiet. He can't force himself to be in an internal state directly (any more than we can *directly* force ourselves to be calm, not be hungry, or not sleepy), and being in certain internal states is simply beyond his control; but he can nevertheless manage his internal states quite a bit.

With a bit of luck, and the contingent fact that Crusoe 4 pretty often has new internal states, he has induced over time a generally positive *success* curve in his interactions with his environment. The same success curve is visible to Crusoe 4 with respect to his applications of his counting words: in the past he often found that he had to go foraging sooner than he expected; sometimes he found himself unexpectedly facing many more predators than he at first thought were there. But over time, as he adopts different methods of counting, these things happen less and less often. Sometimes he can even see how he has induced his success curve to improve over the course of a day.

Notice the idealizations introduced here. First of all, Crusoe 4 experiences as changes in his internal states changes in the counting-algorithms he uses. That's perhaps not so implausible.<sup>1</sup> Accompanying this is his recognition that changes in his behavior can induce changes in his internal states. To maneuver a counting event (or a recognition-of-a-coconut event) so that it takes place in good light rather than in poor light gives him different internal states. A more important (and less natural) idealization is that Crusoe 4's invention of new ways of doing things—counting, in particular—involves no false steps. His new methods are always better methods. This second idealization can be relaxed a little without causing problems for the case; it can be that Crusoe 4 sometimes finds that his new states don't *always* lead to better success in navigating his world; they usually do, however.

God can see how Crusoe 4 keeps inventing better, and then still better, counting algorithms—algorithms that from His point of view enable Crusoe 4 to get the right

---

<sup>1</sup> See, e.g., Strawson (1994), Pitt (2004), as well as the articles in Bayne and Montague (2011), for discussion (and arguments) for there being a phenomenology of cognitive states: a “what it's like” when engaging (for example) in syllogistic reasoning. I'm not arguing for or against this view at the moment. Subject to the discussion of Sect. 3.5, I'm allowing that Crusoe 4 is either aware of the internal states that are his consciously going through this (and not that) sort of reasoning, or that he's aware of internal states correlated with these events of consciousness. Either will do. I should add that I (and others) interpret Wittgenstein as largely denying the existence of this kind of phenomenology. But the absence (or presence) of such a detailed phenomenology isn't (on my view) an *a priori* fact about a being's consciousness of her internal states.

answers (or to get answers closer to the right ones) in successively larger and larger proportions of the counting tasks he undertakes. For example, whereas Crusoe 4 used to count the coconuts in a tree in any old way, now he first sizes up how the coconuts are grouped in the tree, and then first counts the clumps of coconuts in the tree that he's discovered he'll likely overlook otherwise. His counting dispositions have changed over time.

God hasn't a reason, of course, to treat Crusoe 4's words as referring other than as Crusoe 4's (total) dispositions (at a time) incline him to apply them. Because Crusoe 4's meaning dispositions evolve the way they do, God therefore thinks of Crusoe 4 as developing, over time, a continuously-changing *series* of differing disposition-meaning languages, where the words of the later languages fit reality better—or largely fit it better—than earlier ones did. (God can talk this way because He can *see* to what extent Crusoe 4's changing dispositions match the world.)

As the earlier Crusoes did, Crusoe 4 has named collections of objects that strike him as similar. But he's not in a position to easily compare differing disposition-meaning *languages* in terms of their success, as Crusoe 2 does, because Crusoe 4's dispositions are continuously changing over time—his internal states change *a lot*. Crusoe 2 can individuate his (meaningful) words according to which of the two differing meaning-disposition states he finds himself in. Crusoe 4, instead, has to think of himself as speaking *one* language, although the words of that language *wax and wane*—over time—in what they refer to. What they refer to, at a time, is given by his dispositions at that time, and that can change. If he thought to ask himself: “Am I still using these words as I earlier intended?” his answer would be: “Yes I am because I'm, of course, using them as my current meaning-urges incline me to use them.” So Crusoe 4 still understands his language to be a disposition-meaning language.

As a result, Crusoe 4 can't individuate his words by saying (as Crusoe 2 could) that: “I used the wrong word ‘coconut’.” He *can* say “I used the word ‘coconut’ with the wrong dispositions,” because he doesn't allow that his earlier *dispositions* to use his words indicate what his (current uses of his) words refer to. In saying this, Crusoe 4 acknowledges that his current use of “coconut” refers to what his current dispositions impel it to mean, and he's faulting his earlier uses of the word along with his earlier dispositions.

Crusoe 4, however, realizes that the dictates of his current dispositions will be discarded by him tomorrow if his future dispositions then disagree with his current dispositions. How is he to say this? Perhaps by saying, “‘Coconut’ refers to this item today, but it might not tomorrow because I might experience a better internal state when applying the word than I have now.” I won't pause to analyze in detail how this usage of Crusoe 4's is supposed to work. In any case, something like this is reasonable for him to want to say (to himself) because he recognizes that the references of his words will continue to wax and wane in the future: they will always refer at the moment of utterance to whatever his current dispositions dictate them as referring to.

Crucial to the coherence of Crusoe 4's private-language practice is the generally-positive *success* curve that he can induce in his interactions with the environment by *changing* his dispositions in one way, and not in another. Notice that, as I've imagined Crusoe 4, he doesn't think that the range of his words “match up” with the way

natural kinds are really configured; relatedly, he doesn't try to provide an explanation for his success in metaphysical terms. He has only adopted the policy of maneuvering himself to use his words when he's experiencing a new internal state because that's the best way for him to induce a generally-positive *success* curve. In this sense, the meanings of Crusoe 4's words can be understood as genuinely being guided by his intentions. He can be understood as making rational choices because of his ability to remember his previous internal states, and what kind of success he experienced when in those states. He can try to induce new internal states—and he has some ways of doing this. He also has a general policy of trying only to make decisions and take actions when he's experiencing new internal states, or ones that he's found to be pretty good at giving him success in the world.

Two important points should be made now about the preceding discussions of the Crusoe cases. The first is this. I've placed my description of the interactional successes of all the Crusoes against a neat and simple background metaphysical context. I've even given an example where Crusoe 4's generally-positive success curve amounts to what we can call a generally-positive *learning* curve because, metaphysically speaking, his series of disposition-meaning languages have successive words that correspond better and better with how things group into natural kinds in reality.

It's important to realize, however, that such background metaphysical assumptions aren't crucial to the *coherence* of the various private-language practices that these Crusoes have. Describing the various Crusoes' private-language practices as "coherent" means two things: First, such private-language practices holistically contribute positively to a Crusoe's well-being. And second, where needed, a Crusoe has ways of evaluating better and worse alternatives among those private language practices.<sup>2</sup> What's crucial to the coherence of a Crusoe's private-language practice is that he has the capacity to evaluate the success of alternative approaches when he engages with his world, so that he can induce a generally-positive success curve.

The reason that the background metaphysics I've presumed in these Crusoe thought experiments isn't a necessary component of them is because a Crusoe can experience a generally-positive success curve for *all sorts* of metaphysical reasons. It needn't be, for example, that a Crusoe's successive series of languages are (from God's point of view) asymptotically approximating the natural kinds in the world. Clearly, *some* sort of positive dialectical interaction is needed, between the ways the world is and the ways that a Crusoe can change his dispositions, in order for that Crusoe to induce a generally-positive success curve; but this hardly has to involve the kind of simple and clean metaphysical interactions that I've been presupposing. Perhaps, for example, there are no metaphysical kinds at all, and perhaps a Crusoe is protected from this interfering with his inducing a generally-positive success curve because of the strictly local and restricted nature of his interactions with the things in his world. It would be hard to argue that the history of science up until this very day *doesn't* exhibit a pattern like this as opposed to one in which the sciences are progressively converging on some final set of metaphysical kinds.

---

<sup>2</sup> Ways of so evaluating alternative private-language practices aren't needed by Crusoe 1 of Chap. 2, of course, because his experiences of success are ones he can't improve on.

There is a style of philosopher who feels—right here—that a kind of *inference* is called for because a kind of *explanation* is called for. Success, in particular, must be explained by a progressively better matching between our terms and the world.<sup>3</sup> Imagine, for example, that Crusoe 4 has a theory—rudimentary but genuine—of coconuts, squirrels, and so on. His success (on this view) is to be *explained* by the progressively better theories he adopts, and the terms of which better match the world. This terribly seductive “inference” must be resisted.

Epistemological *cul de sacs* are *always* possible, all things being equal. *Mere* success does not all by itself rule out that the success so far enjoyed is only metaphysically local. Nothing rules out the possibility that even a success for all time means only that we are in an epistemic *cul de sac* for all time. This is simply a fact—one that’s often expressed in terms of the limitations of *induction*.<sup>4</sup> Inductions, that is, can always fail—no matter how successful they’ve been (and no matter for how long).

These points are especially important to stress for another reason. As I’ve argued in Chap. 3, a Crusoe can’t judge success by metaphysical comparisons of any sort. What this shows, in turn, is that a Crusoe’s being able to induce a generally-positive success curve is itself necessary and sufficient for his private-language practice being coherent. In what follows, therefore, when I assume that a Crusoe has a successful private-language practice, I’m (usually) just assuming that he has figured out ways to induce a generally-positive success curve in the series of private languages that he invents.

The second point is a caveat to my first point. I’ve described Crusoe 4 as enjoying a generally-positive success curve with respect to his interactions with his environment. But a generally-positive success curve is compatible with things going badly for Crusoe 4 for quite a while. He can go through a period where his current judgment is disturbed, he realizes that, and so he hides out (however well he can) until he’s in an internal state that’s—more or less—reliable. Consider, for example, a Crusoe with a combination of the dispositions of Crusoe 2 and Crusoe 4. This Crusoe’s success curve can shift in a negative direction, when he’s tired for example, or drunk. As long as he has *some* idea of when he’s tired or of when he’s “off his game,” his private-language practice will still be coherent. For although he can’t let the meanings of his words be dictated by the *best* of his previous dispositions (because he doesn’t *have* them at the current moment), he can recognize that those dispositions were better or safer for him and he can “wait it out” until he’s returned to one of those states—until he’s become rested, or sober. (He can also endeavor to remember where his best dispositions agreed with his current dispositions—although, of course, memory and judgment can be affected by these dangerous states.) The crucial element for his private-language practice being coherent is that he can classify better and worse meaning-urges in terms of their impact on his success curve, and he can recognize them introspectively (e.g., feeling tired or dizzy or sloppy).

It’s an empirical question exactly what sorts of evolving internal states are compatible with a coherent private-language practice, and which ones aren’t. If a

<sup>3</sup>Boyd (1991), Putnam (1978, 21).

<sup>4</sup>And so, the insight described here is to be credited to Hume (1961).

Crusoe's powers are precipitously declining, because of Alzheimer's disease, a coherent private-language practice won't be possible. On the other hand, it could just be that a Crusoe's dispositions operate in an arbitrary way that he can't get a grip on at all. Here too, a coherent private-language practice is just impossible. I'm not (nor should I even try to be) prepared to describe specific conditions on the developmental processes of the dispositions of isolated individuals that allow a coherent private-language practice. One reason for this is that what such are isn't just a matter of the individual but of him *and* his environment. I'll describe any set of (changing) dispositions of a Crusoe (in an implicitly-given background world) as a set of "private-language-practice coherence-inducing" dispositions (or, a set of *plpci dispositions*) if they are due to a positive success curve that the Crusoe recognizes and induces. As just noted, whether an individual has such a set of plpci dispositions or not turns not only on him but also on what kind of world he's living in. And, as we've seen, the kind of *world* he's living in isn't something the Crusoe in question can independently assess by comparing it to his dispositions. He can only recognize that the changes in his dispositions that he's induced contribute to an overall positive success curve in his interactions with that world.

## 5.2 A Robinson Crusoe with Plpci Dispositions That He Has—at Best—Inferential Access to

Crusoe 2 and Crusoe 4 are psychologically unrealistic because I've given them what can be called a nearly perfect introspective access to their own dispositions to use words and apply concepts.<sup>5</sup> It's hard to determine a psychologically realistic degree of access of a standard human to her own dispositions to use words and apply concepts (to what degree, with respect to what sorts of dispositions, and how). Finessing that as best I can, though, I'd like, however, to now consider a Crusoe 5 with—largely—*no* introspective access to these dispositions: his dispositions to apply most words and concepts are entirely subpersonal. That doesn't mean, of course, that he doesn't have access to many of his internal states; it's that, for the most part, he's not particularly aware of how these internal states bear on the meaning-urges he has.

In this respect, Crusoe 5's experience of success in the world is quite similar to someone's experience of practicing to shoot an arrow at a target. Among the ways such a person is changing, is that numerous neurophysiological shifts in her are occurring as she gets better at archery. She's not aware of these shifts: she's only aware that she's getting better and better at hitting the target. Of course, she may be aware of certain tricks she's adopted, ways of taking into account external factors (like wind); or she may not.

---

<sup>5</sup>More accurately—and following the discussion of Sect. 3.5 more closely—I've given Crusoe 2 and 4 a nearly perfect correlation between their internal states and their internal-state concepts. I'll generally describe the existence of such correlations as "introspective access" or "awareness."

This is *not* to claim that all of Crusoe 5's psychological states are introspectively unavailable to him (according to God's standards). That—in my view—*would be* psychologically unrealistic. Crusoe 5 certainly has access to those psychological states that—as described in Sect. 3.5—are needed for him to track his success. And so, he has words, and concepts—corresponding to these dispositions that he has access to, ones like “in pain,” and so on. Let's also imagine that, given the world that he lives in, Crusoe 5 has a set of plpci dispositions. Because he has only indirect access to his own dispositions to apply (most) words, I want to describe a very different way that he must think of the meanings of the words of his private language—a way (I claim) that's quite similar to how *we* think about the meanings of the words of *our* (public) language.

Let's start this way. Since Crusoe 5 is unaware of his dispositions to apply (most) words, he's largely only conscious of apparent changes in objects in the world. Objects in the world are sometimes one way to him and sometimes another. I can put it this way: Sometimes they *present* themselves one way to him, and sometimes another; sometimes they present themselves to Crusoe 5 as having certain powers, and sometimes they present themselves as having other powers. Sometimes, when he relies on how objects have presented themselves to him, he doesn't enjoy success; other times he does.

Because Crusoe 5 hasn't introspective access to his own dispositions to apply (most) words, he must recognize indirectly (or infer) any changes in his dispositions to apply those words by how changes in his dispositions have affected his interactions with objects. As already stressed, Crusoe 5 *does* have introspective access to certain psychological *states* he can be in. Apart from recognizing that he's in pain, or afraid, and so on, Crusoe 5 can also recognize that he has become tired, drunk, or dizzy. And he's capable of inferring that his dispositions to apply words have changed because he's tired, drunk, dizzy, afraid, or in pain. But—as my use of the word “inferring” suggests—he recognizes this *indirectly*, because he has learned from previous experience that when he's in such states things present themselves in certain typically different ways from how they otherwise present themselves to him. Things are fuzzy, for example, or more difficult to handle, and so on.

A neat and effective method that Crusoe 5 has to categorize the ways that objects in the world present themselves to him is to distinguish between what an item *looks* like and what it *is* like. The first question to ask about this distinction is *how* Crusoe 5 applies it to *objects*. Note again (from the discussion of Chap. 4) that *no* Crusoe can judge the adequacy of his words by comparing them to the world. Therefore, the distinction Crusoe 5 makes between how objects are and how they seem to be can't be a distinction he draws by comparing how things appear to him to how they are.

How, then, *does* he draw this distinction? Let's first notice that he first learns that objects can shift in all sorts of ways. They are fuzzy, when he presses on the sides of his eyes; they have different colors in different light; and so on. The next question is how he can sort the various ways that he experiences how objects can be into “how they only appear” and “how they are.” The full story, of course, is extremely



complicated<sup>6</sup>; but my purpose here isn't to even try to give a full and accurate story. My only aim is to give an *illustrative* story that shows that there is a way that Crusoe 5's verbal distinctions, between how things appear and how they are, his concomitant notion that he can be mistaken about how they are (because they appear differently from how they are), and his practice of correcting himself (later) in terms of this distinction, can be coherent.

My answer should not surprise anyone who has read this far: as noted, Crusoe 5 has *plpci* dispositions. And, generally, he describes how an object *is*—as opposed to how it appears—in terms of how he “corrects” himself later, when he sides with his currently better meaning-urges. By his “currently better” meaning-urges, I don't necessarily mean the meaning-urges he has *now* (for he may *now* be tired, recognize it, and therefore distrust the way things seem to him to be *now*—that is, he may distrust the dictates of his current dispositions). I mean, rather, those meaning-urges (at any time) that occur during his internal states, and that seem to him to be the most optimal.

An important point to make is that Crusoe 5's distinction between how things appear and how they actually are isn't a systematic one that he can learn to draw all at once. It's a distinction that Crusoe 5 takes himself to be fallible about. *Any* way that he currently thinks objects to be is one that he knows he can subsequently recognize he's mistaken about; so too, *any* way that he thinks objects (only) appear to be may be a way that he can subsequently also recognize that he's mistaken about.

As I've been indicating, this appearance/reality distinction is the basis of Crusoe 5's practice of describing himself as having made a mistake. His subsequent corrections of himself are accompanied by his describing his earlier pronouncements as “mistaken,” in comparison to his current pronouncements. And, of course, the coherence of Crusoe 5's so distinguishing between his making a mistake, and—more importantly—his correcting such mistakes, turns on his dispositions being *plpci* ones.

This isn't the place to dwell on these points in detail, or to refine them, so that the ways of speaking and thinking about mistakes that I've attributed to Crusoe 5 are more accurate to the appearance/reality distinction as *we* actually draw it. Given my description, however, a characterization of Crusoe 5's application of the appearance/reality distinction to the meanings of his words is easy—where this distinction is operating against a background of mistake/correction practices. He takes himself to be applying words to collections of things; he takes himself to be naming kinds of things. Given that he already accepts that something can appear to be different than it is (and that he can subsequently correct himself on this), it easily follows that he can think it's possible for him to believe that something is a coconut because it *appears* to be a coconut, but actually isn't. In correcting himself, he takes himself to have now discovered the object to be a yuckonut instead. He now *knows* the object is a yuckonut although it earlier appeared to him to be a coconut.

Crucial, I claim, to Crusoe 5 being able to apply this kind of talk (about how things appear and how they are) to the meanings of his words is that he's only indirectly aware

---

<sup>6</sup>No doubt it differs, for example, for the different senses, and it involves various sorts of automatic subpersonal faculties.

of his dispositions. He would not even *come up* with this kind of talk if he were directly (introspectively) aware of his dispositions all the time—that he was disposed to have such-and-such experiences when interacting with the world in this way or in that way. Instead he would be focused (as Crusoe 2 and Crusoe 4 are) on which changes in his meaning dispositions induce the steepest positive success curve, and which don't. This isn't an option for Crusoe 5 because he isn't aware of changes in his meaning dispositions "in action," as it were. He can only think about them afterwards—in a more "philosophical frame of mind"—as what he theoretically posits as among the causes of why he thinks something is a coconut, when he does so. In a more searching vein of thought, he can recognize that the ways he categorizes things can only be due to his "dispositions," and not to "the ways that the world is." At that point he will be on his way to discovering Kripke's Wittgensteinian rule-following paradox.

Let's stay focused, however, on a less philosophically-minded Crusoe 5. As Crusoe 5 thinks about his own practices, he sees himself as using words to group collections of objects by their similarities; he thinks that he sometimes gets these similarities right, and that sometimes he gets them wrong. Furthermore, instead of thinking that his success is leading him to group items differently than he did earlier (to ensure that his success continues), he "puts the cart before the horse": he instead thinks that his success turns on his grouping objects correctly—as *they are in the world*. He's consequently prone to speak (to himself) this way: "I'd better get it right that the item up there is a coconut before I climb this tree." He thinks of himself as sometimes wrong, and sometimes right, *about whether something is a coconut*.

Because he speaks of his making "mistakes" in this way, it's natural for Crusoe 5 to take his invented words, "coconut," "yuckonut," etc., to pick out various kinds of objects, and his invented words, "1," "2," etc., to pick out various cardinalities of collections of objects. When he hasn't made any mistakes, that's when he can pick these things out correctly, and that's when he correctly identifies what these words actually refer to. He, of course, thinks of his dispositions to use words—when he thinks of his dispositions at all—as entirely *defeasible*: at any time, he might get it wrong. He also thinks he has some idea of when he's worse at recognizing coconuts (or numbers of coconuts) than he is at other times; and he notices that he gets better at these tasks when he practices.

From God's point of view, what Crusoe 5's "practicing" comes to is this. Crusoe 5 changes his ways of going about his tasks—sometimes in small ways, sometimes in large ways—and he recognizes (sometimes consciously and sometimes subpersonally) when these changes lead to a positive change in his success curve. This is how (again from God's point of view) he's able to develop a set of plpci dispositions over time. That is, most of the time, as he experiments with modifications in how he engages in various tasks, he has more than one option in the directions he can develop his dispositions. If he hits on developing them in the right way, the resulting changes in his dispositions will exhibit a positive movement in his success curve that he can recognize; otherwise they won't. When I describe him as having several options, I mean something like this: He can count carefully or quickly; he can count the items in small groups, and total them, or he can count them in any which way. He can do the task carefully and slowly, or he can rush it. And so on. Thus "what

Crusoe does” must be understood in a very fine-grained way—what’s relevant is that the changes in his behavior affect his answers. And, what causes Crusoe 5 to shift in what he does is whether a positive success curve is induced by one option that’s steeper (in some ways) than the positive success curve induced by other options.

I’ve described all this from God’s point of view. Crusoe 5’s point of view is different: he’s getting better at recognizing various kinds of objects, and he’s getting better at undertaking various tasks: counting coconuts, for example. As he practices, he makes fewer “mistakes.”

Notice that Crusoe 5’s use of normative expressions (“correcting his mistakes,” “being right,” and “doing a task correctly”) is coherent, and coheres with his private-language practice only because he has *plpci* dispositions. He corrects himself, and unbeknownst to him, it’s useful for him to do this *because* the pattern of his corrections (the changes in his current dispositions) enables a positive success curve.

Crusoe 5, as I’ve indicated all along, labors under a serious misconception about how he applies his own words to items in the world. According to God, his words and their references wax and wane over time much as Crusoe 4’s did, although Crusoe 5 doesn’t recognize this. Instead, he thinks his words are fixed in what they refer to, and he thinks his dispositions (at least in principle) can always mislead him. As a result, as I’ve also indicated, Crusoe 5’s beliefs about the words of his private language are pretty much the same as ours (about our public language). Of course, Crusoe 5 thinks he *understands* what his words mean. He thinks, that is, that he can recognize collections of things that are alike (despite his tendencies to mistakes), and he thinks that he knows what functions his arithmetical inventions pick out. Where he has hit on rules for computation, he thinks he’s following those rules when making calculations. He thinks these functions are *not* defined in terms of his dispositions (when he thinks of himself as having dispositions), and he justifies himself in his belief that he’s applying the words as he intended earlier by his confidence in his abilities—always subject to correction by himself at later times. And indeed, he certainly can correct himself (later) when he (earlier) has had too much to drink; he may even, if he tries to calculate during a drinking bout, say to himself: “None of this is right, probably. I’ll straighten it out after I’ve slept it off.”

As I’ve indicated earlier, Crusoe 5 is likely to remain complacent about these beliefs about his private-language practice unless his thinking takes an unusual (a philosophical) turn. Only then is he apt to puzzle himself with questions like: “How do I know the difference between my *really* going on in the same way, and my *only thinking* that I am?” and “What can it possibly be that determines what ‘coconut’ refers to, that’s independent of my dispositions to apply ‘coconut’?”

Let me stress one last time an important reversal in perspective that I’ve introduced in this section, and that’s crucial both to the coherence of Crusoe 5’s private-language practice, and to his nevertheless being mistaken about how that private-language practice actually works. Crusoe 5 thinks he’s successful *because* he’s getting better at recognizing which objects really are similar to which. But this has the situation *backwards*: What’s actually the case is that his dispositions to group things as similar and different are changing over time—although their doing so is largely subpersonal. And what’s guiding this process—and making it not an

arbitrary matter—is that the changes in his dispositions, that he describes as corrections of his earlier mistakes, are induced because they *lead* to greater success in his interactions with the items in his world.

It has often been said that what makes private-language practices incoherent is that because a Crusoe is isolated, whatever is going to seem right to him *is* right (Wittgenstein (1958, § 258)). One reason this has been thought to follow from the sheer fact that a Crusoe is on his own, is that it has seemed that in order to deny that whatever is going to seem right to a Crusoe is right requires that there *be* a standard that a Crusoe's dispositions to use a word must be compared to, and that Crusoe himself must be able to compare his own dispositions to use a word to that standard. More simply put: there must be an external standard of some sort, and this external standard must be one that a Crusoe can *use*.

But this *isn't* needed. *All* that's required is that, first, a Crusoe *understand* any of his applications of his words to be defeasible—that he believe it to be possible that, at any time, he might be “wrong” in his application of one of his own words. What's required, second, is that a Crusoe *have* a coherent practice of correcting himself over time. That is to say, what's required is that his practice of correcting himself not be arbitrary, erratic, or random, and further, that there be a genuine utility for a Crusoe to so correct himself (over time).

What I take myself to have shown in this and the previous sections of this chapter is, first, that if a Crusoe has certain sorts of dispositions, certain abilities to change his dispositions, and a capacity to evaluate the impact of those changes on how well he can navigate his world, there will be a utility in his so-changing his dispositions in order to better navigate his world. Second, if his dispositions are introspectively inaccessible to him, he will be compelled to describe this practice, not in terms of “changing” his dispositions to better navigate his world, but rather in terms of “correcting” his false starts and wrong moves in how he categorizes the items in his world. He will instead have to describe himself as learning to better recognize the things in his world, and how they are similar and different from one another.

### 5.3 Being Guided by Rules Rather Than Compelled by Them

There is a sense in which Crusoe 5 is as good as the Crusoes are going to get: he's indistinguishable from us—in the relevant respects. In order to establish this (in the next chapter), I'm going to turn to other aspects of his language. I'm going to show that his language doesn't need to be analyzed (either by him or by outsiders) in terms of assertability conditions but instead in truth-conditional terms. Further (and this involves the promissory note I made earlier), the semantics that he can attribute (apply) to his language is a standard truth-conditional one.

In the remainder of this section, however, I want to announce a little more loudly an important point about the phenomenology of following rules—about how the

experience of Crusoe 5 when following counting rules compares to our experience of following counting rules. This is important because a major datum that Kripke raises, and that I discussed in the first chapter, is our sensation that we are guided by the rules, rather than compelled by them.<sup>7</sup> I claim that this phenomenological fact is directly due to the dispositions being subpersonal that actually generate the meaning-urges behind counting behavior. Because Crusoe 5's access to his dispositions to use words is like our access to our dispositions, he too feels guided rather than compelled. When someone's dispositions are subpersonal, she can only focus on items in the world (and not on anything internal to her—her reactive dispositions to those items, for example). She can only be aware of how she's appearing to be correcting her earlier impressions of those objects. This is where the phenomenological impression of being guided by one's own concepts comes from. Anyone would have a very different perspective if she were aware of her dispositions, and how she was endeavoring to change *them* (or have *them* change) in light of whether doing so or not would induce a more positive success curve. It's only in these ways that Crusoe 4's impressions of the relationships between his dispositions, his words, and the world differs from Crusoe 5's impressions of these relationships.

Recall the example from Sect. 2.5 where an individual, when counting, experiences a bright visualization of the appropriate numeral that he then has an urge to utter. This is to feel the subpersonal urge that is, in fact, invisible to us humans when we have acquired an ability to count. Notice: the urge needn't be *utterly* compelling; imagine, in fact, that it can be easily resisted by saying something else. (So "compulsion" is too melodramatic a description.) Nevertheless, no one would describe this as being "guided" by rules that had been acquired but (at best) as an urge either to be given into or resisted, like the case of feeling hunger or an itch. In these cases, although the conscious need to eat can often be resisted, "being compelled" or "being pushed" fits the phenomenology far better than "being guided." No one would want to say that she's being guided to eat by feelings of hunger.

## 5.4 Is It Misleading to Describe the Standards for Crusoe 5's Words as "External Ones"?

Are we to think that Crusoe 5 is in *error*, when he thinks he's using external standards for his words? Not necessarily. On one view, he *is* mistaken. He thinks his words match the world and that's where the standards are coming from, that the standards for his correct use of "coconut" are the coconuts. *Wrong*, this sort of error theorist says. The standards are coming from Crusoe 5's own dispositions and from the fact that those dispositions are shifting in such a way as to induce the appearance that the standards are given by the worldly boundaries among kinds of objects.

---

<sup>7</sup> Kripke (1982, 17) writes: "Normally, when we consider a mathematical rule such as addition, we think of ourselves as guided in our application of it to each new instance." Also Kripke (1982, 10): "I follow directions ..."

According to the other view, Crusoe 5 is *right*. Crusoe 5 doesn't understand the machinery of how reference works, but that's a small matter, all things considered. Nevertheless, it's his success over time *in the world* that's inducing the standards he has, and so in this way the world is (globally) inducing a set of standards for his use of words. This *isn't* reference magnetism, of course—not any of the three versions of it discussed in Chap. 3. For one thing, his success could be metaphysically local—a special case that will be discarded if and only if his reach into the world appropriately expands.<sup>8</sup> But nevertheless, the world—or more accurately, his success in his interaction with the world—is playing a major role here in inducing plpci dispositions, and that fact externalizes the standards to the world in a way that's clear. (Certain pragmatists, I imagine, would be very pleased with this second version of the Crusoe 5 story.)

I'll close out this chapter with a brief comparison between Crusoe 5's use of the word "really" and the way that word is used by the earlier Crusoes. The earlier Crusoes all take themselves to be speaking disposition-meaning languages: what "coconut" applies to is whatever their current meaning-urges apply that word to. To ask "is that really a coconut?" allows only the trivial answers, "Yes," or "No," depending on the dictates of current meaning-urges. Crusoe 5 has introduced a new use for "really" (or perhaps a new word, "really") that's based on his distinction between appearances and reality. He can ask "Is that *really* a coconut?" because his current meaning-urges (according to him) don't dictate, for him, that it's a coconut.

For God, of course, Crusoe 5's word "really" waxes and wanes just like his other words. So Crusoe 5 is wrong when he says "That might not be a coconut," or, when he's drunk, "It's not a coconut because earlier I saw it was a yuckonut." "Really" can't do anything more than be dictated to by Crusoe 5's current dispositions. On the other view, he's right because he can allow his word "really" to take account of his best dispositions and not merely his current ones. On this view, he's not speaking the same kind of meaning-disposition language the earlier Crusoes speak. I'll explore the difference between these two views further in Chap. 7, specifically in Sect. 7.2 and 7.3.

---

<sup>8</sup> Similar, perhaps, to how the categories of (Aristotelean) physics were eventually replaced once we recognized friction-tainted environments to be special cases.

## Chapter 6

# Truth and Falsity Attributions and Truth-Conditional Semantics in Private Languages

**Abstract** I continue to explore the language of Crusoe 5, an isolated rule-follower. I show that his language, just like our language, can sustain the use of truth and falsity attributions. That is, it is useful to Crusoe 5, just as it is useful to us, to be able to say that certain groups of statements (or utterances) are true or false. Furthermore, if Crusoe 5 is so inclined, he can develop a truth-conditional semantics for his own language. This is the case even though the sentences of his language are not all underwritten by correspondence facts. These points apply to the natural languages we speak as well.

### 6.1 How Crusoe 5 Speaks of the Truth and Falsity of His Own Statements and How He Talks About What His Words Refer to

As we've seen, Crusoe 5 has a coherent practice of talking about tigers, coconuts, squirrels, and the other things in his world. He can say, that is, "there are three tigers over there" or "there are six coconuts in that tree." But he can say more. He can say, "It's true that there are three tigers over there" or "It's false that there are seven coconuts in that tree." He can also say that he "thought there were three tigers over there, but that's false—it's false that there are three tigers over there. Instead what's true is that there are *four* tigers over there." Crusoe 5 can help himself to talk of truth and falsity and use it pretty much the way we do because he has in place a coherent mistake-and-correction practice of applying his words (one supported by his plpci dispositions).

The uses of the truth and falsity that I've so far allowed him are "redundant" ones, however. They are uses that can be eliminated by putting what needs to be said in a way that leaves out truth and falsity altogether. Crusoe can say, for example, "There aren't three tigers over there," instead of "It's not true that there are three tigers over there." Truth isn't essential to express what he means in this case. We have uses for truth talk that aren't redundant—as when we say that, "Everything Einstein said about General Relativity yesterday is true." This isn't a redundant use of truth because we can't replace that sentence with one in which talk of truth (in some form) doesn't appear if we don't remember precisely everything Einstein said



yesterday. If we *do* remember everything Einstein said yesterday, and that it was only one thing, “ $E = p^2c^2 + m^2c^4$ ,” we can say, “Einstein said ‘ $E = p^2c^2 + m^2c^4$ ’ and  $E = p^2c^2 + m^2c^4$ .”

Crusoe 5 needs these non-redundant uses of “true,” as it turns out, as much as we do, and for the same reasons. He can say proudly to himself, for example, “a number of things that I said yesterday are true.” If he doesn’t remember precisely what those various things were, then his use of “true” in this case isn’t redundant because he can’t eliminate it just as we couldn’t in the same circumstances. These aren’t unimportant non-redundant uses of “true”: Crusoe 5, to the extent that he can, may need to evaluate his own reliability when in certain states. “I often think things that aren’t true when I’m drunk,” he may conclude, for example; and this thought—couched in a non-redundant use of the truth—may be extremely important to him. This is especially the case because—like us—he often remembers the general contours of an event (like: he got a lot wrong) while failing to retain the particulars (*exactly* what he was wrong about—*which* sentences).

All of this suffices for Crusoe 5 to have a full and robust truth notion—at least according to the various philosophers who are truth deflationists.<sup>1</sup> That is, his self-attributions of truth and falsity are coherent because (i) they’re based on an antecedently coherent mistake-and-correction practice applied to his sentences, and also because (ii) his truth and falsity phrases are linked to his other sentences by obeying the Tarski biconditionals—all sentences of the form,

“There are three coconuts in that tree,” is true if and only if there are three coconuts in that tree.

Crusoe 5 talks (to himself) about truth and falsity in just the way *we* talk to one another about truth and falsity (and sometimes to ourselves as well). But even further (and just like us), Crusoe 5 also talks about his words *referring* to things or to classes of things in the world. And similarly to his self-attributions of truth and falsity, he takes it that he can be wrong about what he thinks he’s referring to with a word. He can be wrong about *that* being a coconut, or he can be wrong about coconuts all being edible. Because he thinks he can be wrong, he thinks that the standards for whether he’s right (or wrong) about whether something is a coconut or not is whether or not the thing in question *really is* a coconut.

Crusoe 5 takes his word “coconut” to refer to all and only the coconuts. He takes his word “five” to refer to all and only collections of five objects. He talks about correctly seeing that something is a coconut, and sometimes incorrectly thinking something was a coconut. He talks about making mistakes when he counts coconuts, and correcting those mistakes later. It’s the positive success curve, that the changes in his dispositions induce, that makes this practice coherent—along with his talk of truth, falsity, and reference.

God’s picture of Crusoe 5’s private-language practice is (unsurprisingly) very different from Crusoe 5’s picture of it. From God’s point of view, as I described it in

<sup>1</sup>There are many kinds of truth deflationists; for needed distinctions among them, see Azzouni (forthcoming). For arguments for my favorite brand of the position, see Azzouni (2006, 2010b).

Chap. 5, Crusoe 5 doesn't have a single stable language; he has a continually-changing series of disposition-meaning languages. Or, equivalently, Crusoe 5's words keep shifting in what they refer to. (The *references* of Crusoe 5's words *wax and wane* over time.) From God's point of view, Crusoe 5's words refer (at a time) only as Crusoe 5's best dispositions (at that time) dictate them to refer. But the extensions of Crusoe 5's words keep changing because of the positive learning curve he induces in his word-application practices. Given *one* possible background metaphysics, God sees why Crusoe 5's success in his world keeps increasing: the extensions of his words more and more closely approximate the natural kinds that there are (that God has created). Given a rather different background metaphysics—because God, after all, has a lot of options about this—God sees Crusoe 5's success much more ironically: it's due to an epistemic *cul de sac*. Crusoe 5 in this second case is enjoying what God recognizes to be a strictly local (and perhaps temporary) success.

Imagine a case where Crusoe 5 has a set of dispositions that—at one point—dictate no answer to the question: how many coconuts are there in that tree? There are too many coconuts in that tree for him to manage counting them consistently. After he makes several attempts, he's disposed to recognize that his answers are varying too much, and so he doesn't trust his abilities. Later, his dispositions change, and he can successfully produce a stable answer. He thinks: "only *now* do I know that the number of coconuts in that tree is 57. It always was 57, of course, but I didn't know that until now. It was always *true* that the number of coconuts in that tree was 57 even though I was once incapable of finding that answer out. Perhaps there are trees out there with so many coconuts that it's impossible for me to ever know how many they have. Nevertheless, for each coconut tree—no matter how many coconuts are in it—there is a number such that it's a fact that the coconut tree in question has that number of coconuts."

Was there a fact of the matter that there were 57 coconuts in that tree before Crusoe 5 developed the capacity to give that number as the result of a counting task? In asking this, despite appearances, I'm asking a question *within* Crusoe 5's language, using his word "57" and not my word "57." Putting sentences and words of Crusoe 5's language in c-quotes (to indicate that I'm quoting his language despite its resemblance to mine) resolves any danger of misinterpreting my question, for my question can now be posed this way: Did a fact correspond to 'there are 57 coconuts in that tree'?<sup>2</sup>

I discussed one tempting way to resolve this in Sect. 2.2: describe the fact in question as about how Crusoe 5's dispositions *will* develop (under such and such circumstances). Since he will (someday) have dispositions that dictate '57' as the answer to the number of coconuts in the tree, there is a timeless fact of the matter about this. But Crusoe 5, as we've seen, thinks that there is a definite number of coconuts in *any* coconut tree. He thinks this even about those collections of coconuts that it's impossible for him to *ever* be disposed to count. So, regardless of whether we give in to the

---

<sup>2</sup>This is a grammatically correct question of my language—English—just as "Does 'Chaque champignon est vénéneux' correspond to a fact?" is a grammatically correct question of English. My thanks to Douglas Patterson (9/24/09 email) for his complaints about an earlier formulation.

temptation to characterize the facts about the number of coconuts in possible coconut trees in terms of how Crusoe 5 will be disposed to count, Crusoe 5 clearly takes his words to refer beyond the scope of his dispositions, and he takes his talk of what's true and false to extend beyond the scope of his dispositions as well. For us (and for God), there are no grounding *facts* that underwrite these claims of Crusoe 5's; consequently, there are no correspondence facts that determine their truth.

We (and God) might be tempted to translate Crusoe 5's terms to ours (to His). In particular, we might think it appropriate to translate Crusoe 5's 'number<sup>c</sup>' terms to our "number" terms in order to legitimate the truth value for Crusoe 5's 'there are 57 coconuts in that tree<sup>c</sup>', and other statements of the same form, regardless of whether Crusoe 5's dispositions can *ever* provide grounding facts for them. Translating his words to ours with this motive *faces the original rule-following paradox*. There are no grounds that can be used to legitimate a translation to our "number" as opposed to an alternative "number\*." Furthermore, it's hard to see what (*a priori*) constraints one could (legitimately) place on translation practices that would justify ourselves (or God) in so translating Crusoe 5's words to impose determinate answers to questions that are beyond him. If the original rule-following paradox—as described in Chap. 1—has validity, then it has validity here as well.<sup>3</sup>

Someone who grants the coherence of the private-language practices of the various Crusoes might argue that what I've shown, rather surprisingly, is that assertability conditions don't require a community, as Kripke's Wittgenstein requires them to. Given that an isolated Crusoe 5 has a plpci set of dispositions (whether he realizes this or not), he's capable of a private-language practice governed by assertability conditions. Suppose, for example, that such a Crusoe 5 has acquired dispositions to add (up to a particular number, of course). One might try this as a suitable assertability condition: Crusoe 5 is entitled to claim that he means addition by "plus" (at least up to a certain number) because (roughly) he's confident that he can give correct answers to new cases, subject to corrections by his better self—and where such a better self is characterized in relation to Crusoe 5's plpci dispositions.<sup>4</sup>

I *don't* want to claim this. I claim, instead, that Crusoe 5's language can be regarded as a "truth-conditional language," one pretty much like ours, and one that can be characterized the way ours is. I take this up in the next three sections.<sup>5</sup>

---

<sup>3</sup>Notice that the force of this claim turns (in part) on the rejection of reference magnetism that I give in Chap. 4. This *doesn't* mean, however, that translating Crusoe 5's words to ours is ruled out. It isn't ruled out as long as we don't delude ourselves into thinking we'll solve the rule-following paradox by doing so. I'll address this in 6.4, but I can say this much now: the rule-following paradox applies to us as well as Crusoe 5 (although not to God). Because of this—provided other constraints are in place—we can justify a translation, although not one that provides grounding facts.

<sup>4</sup>I don't think this can be made to work, but I won't pause further to give details now.

<sup>5</sup>For purposes of argument, I'm adopting the position in this book (except for occasional demurals) that truth-conditional semantics is both a useful (and informative) semantics, and one that applies to natural languages. Many linguists (and philosophers familiar with linguistics) will disagree with the second assumption. See the discussion of Chomsky and Pietroski in Sect. 4.4. As far as the first assumption is concerned, see Azzouni (2008) for worries about the possible triviality of (most) truth-conditional semantic analyses of languages.

## 6.2 Truth Conditions, Bivalence, and the Broad Ignorance Thesis—Applied to Our Own Language

One thing that's needed to support the claim that Crusoe 5's language can be treated as a truth-conditional language despite the rule-following paradox is the denial of an absolutely critical assumption of Kripke's Wittgenstein. This assumption, of course, isn't restricted to him, but is widely-held: truth conditions *require* correspondence facts. As various sorts of truth deflationists have argued,<sup>6</sup> however, the sentences of a language can have "truth conditions" even if truth isn't a correspondence notion.<sup>7</sup>

I'm going to develop this point in this and in the next two sections by first using different examples, ones I've used in other work, and then I'm going to apply the results to Crusoe 5. I won't defend my truth deflationism against competing views—I've already done that extensively in other work<sup>8</sup>—I'll simply state the view and sketch some of the arguments for it. Two points about this expository strategy. First, as I just indicated in footnote 8, I've extensively defended my particular truth-deflationist view elsewhere; if the reader needs more details, they're easily available. But, second, I understand Kripke's interpretation of Wittgenstein's strategy to be an "only game in town" move. We're *forced*—or so it looks—to a sceptical sociological solution of the rule-following paradox because there seems to be no other option. As the discussion of reference magnetism in Chap. 4 makes clear, it's likely that Lewis's response both to Kripke's Wittgensteinian rule-following paradox and to Putnam's "paradox" is similar in spirit: this is why Lewis so often speaks of himself as responding to a *reductio*—a rhetorical maneuver imitated by many of his followers. So my aim here isn't so much to argumentatively oppose these two families of alternatives with my particular sceptical solution to the rule-following paradox (although I've engaged in quite a bit of that, nevertheless); I primarily aim to present another cogent position so that these "only game in town" strategies are undercut.

I'll begin by speaking of our own language, not Crusoe 5's. Here are a number of linked claims about our (or, I suspect, any) natural language:

- (i) Natural language contains sentences with non-referring terms.<sup>9</sup>
- (ii) Despite (i), many of those sentences have truth values.

<sup>6</sup> See the citations in footnote 1, both for my particular truth-deflationist position and for discussion of the surrounding truth-deflationist literature. I discuss my truth-deflationist position somewhat briefly in what follows.

<sup>7</sup> Some philosophers—in denying truth to be a correspondence notion—also deny that *any* truths correspond to facts. But we don't have to assume that drastic position as far as Crusoe 5 is concerned. All that's required is that not *all* his truths correspond to facts. This fits well with *my* general deflationist position—as I'll indicate shortly.

<sup>8</sup> Azzouni (2006, 2010b), for example.

<sup>9</sup> As I'm using "refer" here, it characterizes a real metaphysical relation between words and objects, and not a pleonastic one. In this substantial sense of refer, "Mickey Mouse" doesn't refer to Mickey Mouse, because there isn't any such thing. "Refer," so used, isn't natural language (since it's natural in natural language to say that "Mickey Mouse" refers to Mickey Mouse). It's instead a terminological convention that I've introduced. See Azzouni (2012a) for a full discussion of how we

- (iii) Compatibly with (i) and (ii), a portion of natural language can be treated as having a bivalent classical logic.
- (iv) As a result of (iii), a portion of natural language can be treated as possessing a standard truth-conditional semantics.
- (v) Crusoe 5's language is the same as ours in these respects, and so (i)–(iv) hold of his private language as well.

Qualifications (and notice these are substantial): I wrote in (iii), “compatibly with these two previous facts” because I don’t want to commit myself to the claim that it can be established of even small portions of natural language, that they *in fact* obey (one or another) bivalent classical logic. I doubt this, although I think an historical event occurred in the early twentieth century the result of which is that we now impose such a logic normatively on natural languages (at least insofar as they are impounded for use in the sciences). This is a big topic, one that I address elsewhere, and I really can’t say more about it now.<sup>10</sup> For our purposes, the right way to put the point is this: as far as the rule-following paradox is concerned, Kripke’s considerations (on behalf of Wittgenstein) don’t impel the replacing of truth conditions with assertability conditions.

I turn to my sketch of the justification of (i)–(v). The first point to make is this. It’s often *indispensable* to a language practice—private or public—to describe sentences as either true or false, even though some of those sentences have terms that don’t refer.

Consider these examples.

- (1) Minnie Mouse has *never* been depicted in movies as a plumber.
- (2) Cartoon characters and ordinary people exist who look like one another.
- (3) Strategies for circumventing anger exist. Many are found in self-help books.
- (4) There are as many prominent Greek gods as there are prominent Greek goddesses.
- (5) James Bond is depicted in Ian Fleming’s novels as suave and sophisticated although he isn’t depicted this way in recent movies.

All of these sentences have truth values (although I don’t know the truth value of (1)). This *isn’t* because “Mickey Mouse” or “Minnie Mouse” or “James Bond” refer; it isn’t because “cartoon characters”—in the sense meant—or “strategies” or “Greek gods” or “Greek goddesses” have non-empty extensions.<sup>11</sup> The view taken here is that these sentences have truth values despite containing (some) terms that don’t refer to anything, and despite there being no metaphysical facts that these

---

should understand “refer” in natural languages, and how we should refine that notion for philosophical regimentations.

<sup>10</sup> See Azzouni (2006, 2009, 2013), especially the last chapter.

<sup>11</sup> Kripke (2013), famously, has an influential competing view: fictional names operate pretentively (in fictional contexts—novels, plays, and so on) and they also refer to abstracta—“characters.” They have a third use in, for example, negative existentials. Kripke’s view is complex and attributes ambiguity to these sorts of names.

sentences either do or don't characterize correctly (for example, the *ways* that the things are that the terms in these sentences refer to).

As I just suggested, a natural picture of how all this goes is a "truth-maker" view. A sentence is about various objects and it's the ways that those objects are that determines the truth or falsity of the sentences about those objects. If how a sentence describes certain objects as being is how they are, then that sentence is true; otherwise, it's false. When sentences have terms that don't refer, however, this can't be—in general—the right picture of how their truth values are determined. Instead, there is a human practice of some sort in place that assigns truth values to these sentences, where such a practice is sensitive to relevant aspects of the world, and in this way the assigning of these various truth values to sentences is forced.

So, for example, we have a practice of making up fictions. The actual fictions themselves, novels for example, are composed of sentences that don't have truth values—they are *pretended* (by authors and readers) to have truth values. But when we talk *about* the characters in the stories, as we do about "James Bond" in (5) above, the sentences often have specific truth values *induced* by the stories that form the basis of the practice of assigning truth values to these sentences. In the fictional cases, how this occurs is fairly straightforward (although often difficult to precisely determine in practice): a character is depicted in such-and-such ways in a piece of fiction (just as real people can be depicted in such-and-such a way in a piece of fiction or in articles about them) and a true sentence about that character describes the depiction correctly.

An issue is immediately intruding. This is that it's pretty clear that although many sentences, about James Bond for example, are truth-value induced by the stories (or movies) that James Bond appears in, there are infinitely many sentences that aren't truth-value induced by story-telling practices, no matter how extensive those practices are. Some philosophers have quickly drawn the conclusion that the "logic" of fiction can't be a bivalent classical one, but instead must be some gappy alternative. Others have drawn the conclusion that all the sentences that contain non-referring terms must be treated non-literally in some fashion—e.g., as involving semantic pretense, or a kind of fictional operator.

These drastic strategies aren't required, however. We can treat *all* our sentences about a subject area as belonging to two categories: those we *know* the truth values of and those we don't. That is, even if we know that a sentence doesn't have a truth value, we can still speak as if it does, and describe ourselves as ignorant of that truth value. And, actually, this is pretty natural. For example, while reading about something James Bond does, we might idly wonder if we're reading about a genuine sociopath. The correct answer is that there is no answer to this question—not even in principle—because the fictional practice determines facts about a fictional character by the author explicitly stipulating them (in a piece of fiction) or by the author acquiescing in them by working within a particular genre.<sup>12</sup> So, for example, James

---

<sup>12</sup>With some exceptions: for example, facts about whether a character is famous or worshipped, and so on. These are induced by the reactions of actual people *to* the stories or depictions of the characters.

Bond—in the Ian Fleming novels—isn't depicted as an intelligent reptile from outer space. We know this not because there is a sentence in any of Fleming's books that implies this truth but because Ian Fleming just isn't writing that kind of novel. But the question of whether James Bond is being depicted as a sociopath in Fleming's novels is a question that's too subtle either for the author or for the genre to have dictated an answer to. (Literary criticism will grind to an indeterminate halt over this question—if it ever comes up in published discussion.)

We express ignorance about the truth values of sentences about fictional beings that aren't fixed by our fictional practices. As I mentioned, we can say (extremely naturally) that, "I don't know if James Bond is really being depicted as a sociopath—no one does." This expression of ignorance is the same as when a fictional practice does fix the truth value of a sentence, but we happen not to know what it is, for example, "I don't know how many daughters Lear is depicted in Shakespeare's play as having, but I can look it up in Wikipedia if I have to." In point of fact, our verbal practices in many areas of life fail to dictate—even in principle—what truth values certain sentences should have. This is true in mathematics—where proof (except in certain exceptional cases) is out-stripped by truth—and in ordinary life, where nothing (even in principle) dictates whether or not certain people are bald. In these cases, we speak of being ignorant of the facts—regardless of whether, metaphysically, there are facts to be ignorant of.

Some philosophers think this way of speaking needs correction.<sup>13</sup> They argue that the actual logical principles—that we express ourselves as committed to—should respect the metaphysical facts about when statements "really" have truth values. So, in this case, truth-value gaps should be explicitly accounted for by changing the logic to accommodate them, and not by burying those gaps in expressions of ignorance. But why? Officially deserting classical bivalent logic is costly—not least because it best accords with our implicit inference practices.<sup>14</sup>

Accompanying our adoption of a bivalent classical logic is knowledge/ignorance terminology that handles both kinds of truth-value uncertainty. "We don't know," we say, both when we don't know enough to say whether someone is bald and when it's the case that nothing in our (collective) understanding of how the word "bald" applies to heads can determine whether the term "bald" applies or not to a particular head. Furthermore, as the numerous cases of vagueness in ordinary language make clear, there is often no "bright yellow line" marking out those statements that "metaphysically" lack truth values from those that we've only so far failed to determine truth values for. One reason is that revisions in our methods for determining the

---

<sup>13</sup>Priest (2011, 361), with respect to how truth outstrips proof in mathematics, argues that "it will often be the case that we can establish neither  $A$  nor  $\neg A$ . So it would seem that we have truth-value gaps of some kind. How, then, can we justify using classical logic? ... Classical practice establishes everything of the form  $A \vee \neg A$ . So the classical account of disjunction would seem to go."

<sup>14</sup>This is important. As I indicate in the work cited in footnote 10, it isn't that our implicit inference practices actually exemplify one or another instantiation of classical logic. It's rather that our implicit inference practices are best regimented by classical logic. They're best treated as obeying those principles as a norm.



truth values of statements or in how we take our terms to apply may change the class of “determinable” statements.<sup>15</sup>

Naturally, though, some philosophers argue that our ways of self-describing ourselves as ignorant *require* that there *is* a “fact of the matter” of which we are ignorant. If there is no fact of the matter whether the word “bald” applies or doesn’t apply to a particular head, then we aren’t ignorant of whether or not it applies—so these philosophers argue. But everyone *does* speak of being ignorant in just these cases (“I don’t know whether he’s bald or not—*no one* does.”). That is, we’re faced with a situation in which our understanding of the predicate (“bald,” say) doesn’t reveal whether the predicate applies or not. It’s precisely at these times that we speak of being “ignorant.” Notice therefore: this *isn’t* the claim that (unknown to us) there actually *is* always a fact of the matter about whether “bald,” say, applies or doesn’t apply to a case that otherwise strikes us as indeterminate. *That’s* Williamson’s view (1994).

It’s quite natural, apparently, to think that if not every sentence is “determined” as true or false, then one isn’t “entitled to assert that the principle of bivalence is true,” as Walker (1989, 33) writes. We must tread carefully. To say—straight out—that there are sentences that are neither true nor false is, of course, to *officially* embrace a denial of bivalence, or it’s—at least—to use “true” in a way that disallows semantic ascent. That’s why I informally employed the weasel word “determinately.” But to say that there are sentences which *we don’t know* to be true or false isn’t to do either of these things. To say, further, that there are sentences which (most likely) we’ll never know to be true or false isn’t to deny bivalence either—because we’re making a prediction about what we’re never going to “know,” and in particular, we aren’t (on pain of contradiction) connecting the permanent and principled unknowability of a sentence with its “lacking a truth value.” So to say that there is “no fact of the matter” about whether a sentence is true or not is *not* to say that it isn’t the case that the sentence is either true or false. (For to say that *would be* to desert bivalence.)

Think carefully about what we *do* say. We say, for example, “Peter is bald or it isn’t the case that Peter is bald.” Using the Tarski-biconditional property of “is true” (and a couple of other straightforward principles), we can also say, “‘Peter is bald’ is true or ‘Peter is bald’ isn’t true.” Do either of these statements say that the sentence “Peter is bald” has the truth value *true* or the truth value *false*? The first statement is only a disjunction of two statements. There is no denying that a number of inferences are justified by this disjunction. But they all turn on additional facts about the particular sentence “Peter is bald.” As long as we don’t know, in fact, whether Peter is bald or not, we aren’t licensed to draw anything specific *about Peter* from the disjunction. (That’s the sort of thing many people mean when they describe this disjunction as a “tautology.”) The second sentence is similar in this

---

<sup>15</sup>I call this view of the “don’t know” phrase, the “broad ignorance thesis.” See Azzouni (2010b, 91–93). This particular way of handling truth value gaps—that can be labeled “The compatibility of classical logic with de facto truth-value gaps by means of expressions of ignorance”—is one I first offered in Azzouni (2000), Part IV, § 6.

respect: We can't draw any specific conclusion either about Peter or about "Peter is bald," for that matter.

An irate opponent might complain that the second sentence *says* that a certain sentence is either true or it's not true. A disjunctive *property* is being attributed to a sentence. Response: It's similar to saying, "Mickey Mouse is more famous than Hillary Clinton." There is *no* Mickey Mouse—nevertheless the sentence "Mickey Mouse is more famous than Hillary Clinton" is truth-induced by our cartoon practices. "'Peter is bald' is true or it isn't the case that 'Peter is bald' is true" is also truth-induced by our (logical) practices—in lieu of any facts that force one or the other truth value on "Peter is bald." What's the problem, therefore? Just as someone might complain "Mickey Mouse is more famous than Hillary Clinton," *can't* be true *or* false because there is *no* Mickey Mouse, someone might complain that "Peter is bald" *can't* be true *or* false because there are *no* facts to determine its truth value. In both cases, how we talk about the truth and falsity of sentences outstrips the correspondence relations correspondence theorists of truth try to impose on truth-talk. So much the worse for correspondence theorists of truth and their impositions.

I've noted that we can tussle over our talk about *truth values* in exactly the same way that we can tussle over talk about *property attributions to what doesn't exist*. But if we can accept one way of speaking, we can accept the other. We can thus speak of a sentence having the property of being true or not true (even if it's neither—even if *there are no sentences*, as nominalists are likely to claim) just as we can speak of a nonexistent object having a property or as not having a property (even if *it* doesn't exist). Alternative ways of speaking, of course, are ways of deserting standard bivalent discourse (by introducing additional or "determinate" truth values). The suggestion I'm offering is to instead embed this complexity into talk of knowledge and ignorance, and thus leave both our talk of truth and falsity, and the related talk of truth values untouched. I claim further that this is already our practice in natural languages.<sup>16</sup>

### 6.3 Tarski Biconditionals and Truth-Conditional Semantics—Applied to Our Own Language

Standard truth-conditional semantics applied to a language that lacks context-sensitive terms (terms like "that," "he," "I") is supported on a base of a set of Tarski biconditionals. Otherwise (there are two options) either it's also supported on a base of Tarski biconditionals or alternatively it's supported on a base of what are called "conditionalizations." It's commonly believed that truth-conditional semantic theories of all these sorts provide language-world connections. The crucial point here is

---

<sup>16</sup>Recall the concern raised in the general introduction, footnote 4, about Kripke's phrase "meaningful declarative sentences must purport to correspond to facts," and notice that there is a sense in which this can be *accepted*. We can express ignorance about what those facts are even when we (otherwise) think there are no such facts.

similar to the crucial point established in the last section. We saw that talk of the truth and falsity of sentences doesn't require that the terms in those sentences refer to anything in the world—language-world relations needn't be relevant to talk of truth and falsity. So too, a truth-conditional semantics can apply just as well to languages where terms don't refer as it can to languages where the terms do refer—so here too, language-world relations needn't be relevant. Notice the neutrality about these language-world relations that I'm stressing: "needn't" doesn't mean "doesn't."

Suppose a very simple language has one predicate, "is a coconut," and two names, "Julie" and "Mike." Presume that "is a coconut" picks out all and only the coconuts. Presume that "Julie" and "Mike" are specific individual things designated by those names. In addition, the language has one connective, "&," and one quantifier, "there is." Nearly enough, here is a (baby) truth-conditional semantic theory for this (very restricted) language:

"Julie is a coconut" is true iff Julie is a coconut.

"Mike is a coconut" is true iff Mike is a coconut.

"Something is a coconut" is true iff something is a coconut.

If  $S_1$  and  $S_2$  are sentences, then " $S_1$  &  $S_2$ " is true iff  $S_1$  is true and  $S_2$  is true.<sup>17</sup>

Notice that the last clause certainly doesn't "touch base" with the world, as it were. Instead it coordinates the truth values of certain sentences with the truth values of other sentences. The first three clauses, however, may seem different. It's true that they also coordinate the truth values of sentences with other sentences, in particular, the first clause coordinates the truth of "'Julie is a coconut' is true" with "Julie is a coconut." But the sentences on the right sides of these biconditional clauses nevertheless seem to directly describe situations in the world—because they're "used" not "mentioned." Although "'Julie is a coconut' is true" is about a sentence being true, "Julie is a coconut" is about Julie being a coconut. The conclusion would seem to be that therefore such semantic theories necessarily describe language-world relations, and they do so pretty much because they are based on Tarski biconditionals.<sup>18</sup>

Consider, however, a language which allows sentences of the sort I gave in Sect. 6.2. Suppose, in particular, that it has one predicate, "is depicted as a plumber," and two names, "Mickey Mouse" and "Minnie Mouse" (neither of which refer), although presume that "is depicted as a plumber" picks out all and only the things that are depicted as plumbers. In addition, as before, the language has one connective, "&,"

<sup>17</sup> *Some qualifications.* The reader familiar with these sorts of theories will notice a large number of liberties that I've taken. Among them, I've suppressed discussion of the metalanguage/object language distinction, I'm using variables within quotation marks, and I'm burying details about the quantifiers—including issues about satisfaction. All of this can be cleaned up nicely, of course—and it's been done in numerous places, including in many places in my own work, e.g., Azzouni (2008) or Azzouni (2010b).

<sup>18</sup> This view is very widely held, and is seen by many as an advantage of such semantic theories over ones that don't "disquote." I discussed this in some detail in Sect. 4.4. See, for two more examples among numerous others, LePore (1983) or Ludlow (1999).

and one quantifier, “there is.” Nearly enough, here is a second (baby) truth-conditional semantic theory for this (equally restricted) language:

“Mickey Mouse is depicted as a plumber” is true iff Mickey Mouse is depicted as a plumber.  
 “Minnie Mouse is depicted as a plumber” is true iff Minnie Mouse is depicted as a plumber.  
 “Something is depicted as a plumber” is true iff something is depicted as a plumber.  
 If  $S_1$  and  $S_2$  are sentences, then “ $S_1$  &  $S_2$ ” is true iff  $S_1$  is true and  $S_2$  is true.

In this case, just as earlier, we have the appropriate sentence-to-sentence covariations in truth values—but there are no language-world relations *at all* because there is no Mickey and Minnie Mouse. Nevertheless the semantic theory takes exactly the same form.

There may seem to be a very large elephant in the room that I’m not acknowledging. This is that the truth-conditions given just now for these sentences that contain terms that refer to nothing at all seem *not* to supply necessary and sufficient truth conditions, at least in one way that necessary and sufficient truth conditions are understood. The left-hand side of each of the clauses in these truth conditions relates the sentence to an object that doesn’t exist, and the right-hand side of each clause describes the truth condition of the sentence in terms of a condition on that (nonexistent) object. But there are no such objects (by assumption) and so they have no properties. How, then, can these clauses supply truth *conditions*? How can they reveal the circumstances under which these sentences are true or not? And isn’t revealing the circumstances under which sentences are true or not what’s meant by describing such semantic theories as *truth-conditional* semantic theories?

So consider the truth condition:

- (6) “Mickey Mouse is depicted as a plumber” is true iff Mickey Mouse is depicted as a plumber,

How is this truth condition for “Mickey Mouse is depicted as a plumber” supposed to be connected to the *actual truth* of “Mickey Mouse is depicted as a plumber”? How can such a supposed truth condition determine the truth of the sentence within quotations on the left-hand side of (6)?

It can’t—not all by itself, anyway. And here I must delicately take the bull by the horns (or, more accurately, the elephant by its trunk): “Truth conditions” is misleading terminology that was originally used to describe a certain (Tarskian) style of semantic theory because of the false impression that such semantic theories must yield (that is, provide determining conditions for) the truth values of the sentences to which the semantic theories apply. On the view of semantics being argued for here, truth conditions can be given to a set of expressions exactly as it’s done in the various semantic traditions. Furthermore, these conditions are *compatible* with the truth values that these expressions actually have. For example, on this approach, the truth of the sentence “‘Mickey Mouse is depicted as a plumber’ is true” is correlated with the truth of “Mickey Mouse is depicted as a plumber,” even though “Mickey Mouse” doesn’t refer. That’s not a problem since (as established in the last section) the sentences in question can have truth values despite this. A fictional practice, that’s in place with respect to terms like “Mickey Mouse,” induces a truth value for

both these sentences, or neither, equivalently. Traditional truth-conditional theories, therefore, don't give truth conditions that *determine* the truth values of the expressions they're about—under all circumstances. Or rather, they do so in pretty much the same fashion that the following clause provides truth conditions:

- (7) “Napoleon was worshipped by the ancient Greeks” is true if and only if Napoleon was worshipped by the ancient Greeks.

The sentence on the left-hand side of (7) is true if and only if the sentence on the right-hand side of (7) is true. The sentence on the right-hand side isn't true, and thus neither is the sentence on the left-hand side. Exactly the same thing can be said about (6).

Let's now consider sentences of a rather different sort, ones that seem to connect language to the world in a more direct way. I'm thinking here of terms like “that,” as in “that vase,” or “she” when the individual being indicated is in the room, or even cases where a description is given, such as “the table,” but everyone involved knows which table is relevant because there is only one table in the room that anyone can be talking about. Utterances of these sorts of sentences have been handled by interesting extensions of the truth-conditional approach to semantics first pioneered by Tarski. Here too, although it may seem that the semantic theories in question are restricted only to contexts where the uses of such terms are really picking out things in the world, this isn't so. Consider, first, a fairly nontraditional approach to the semantics of the uses of these sorts of terms.<sup>19</sup> In this one, context-sensitive expressions (which are treated as co-referring) appear on both sides of the clauses governing utterances of those expressions. That is, if all the following are utterances, then:

- (i) “That vase is ugly” is true if and only if that vase is ugly,
- (ii) “She is an hallucinated object” is true if and only if she is an hallucinated object,
- (iii) “The hobbit over here isn't real” is true if and only if the hobbit over here isn't real,
- (iv) “I am hungry” is true if and only if I am hungry.

These are taken (on this approach) to provide some of the various truth-conditions that the theory (of this piece of language as a whole) is supposed to yield. Notice that the context of utterance of both sides of these biconditionals is required to be the same. A variant set of truth conditions, ones that take account of the context of the interpreter being different from that of the utterer of the utterances in question, appropriately shifts the context-sensitive expressions like so:

- (iii\*) “The hobbit over here isn't real” is true if and only if the hobbit over there isn't real,
- (iv\*) “I am hungry” is true if and only if he is hungry.

---

<sup>19</sup>I draw this approach from Ludlow (1999), chapter 3.

Providing a truth-conditional interpretation of someone else's utterances (at the moment that person is making them) may acceptably include context-sensitive uses on the right side of the biconditionals that appropriately refer to items in the interpreter's context. The thought is misguided that this sort of semantics *only* works when all the uses of the terms in an utterance refer. On the contrary truth-conditional clauses of either sort allow non-referring uses of singular terms to appear on both sides of the appropriate truth conditions. If one can be in a context where it makes sense to gesture towards one's (or someone else's) hallucinated object, and utter a demonstrative sentence with terms that purport to refer to that hallucinated object then a truth-condition clause can be supplied that avails itself of exactly the same resources. In short, uses of terms on the right side of a truth-condition clause can purport to refer to the same items as uses of terms on the left side of a truth-condition clause, even if the terms don't refer to anything.<sup>20</sup> And contexts—the background from which interpretations for uses of context-sensitive terms are supplied—needn't be restricted to ones where those context-sensitive expressions are used only to refer (to what's real).

As I mentioned, however, the above sketch of an approach to truth-condition clauses, isn't the standard approach for handling context-sensitive expressions. It's more common to express the content of sentences with such context-sensitive expressions by "conditionalizations"—descriptions of their purported referents. Instead of (i), (ii), (iii), (iii\*), (iv) and (iv\*) above, we have:

- (i\*\*) An utterance *u*, at time *t*, by speaker *s*, of "That vase is ugly" is true if and only if there is an object (a vase) *o*, designated by *s* by her use of "That vase" at time *t*, and *o* is ugly.
- (ii\*\*) An utterance *u*, at time *t*, by speaker *s*, of "She is an hallucinated object" is true if and only if there is an object *o* designated by *s* by her use of "She", at time *t*, and *o* is an hallucinated object.<sup>21</sup>
- (iii\*\*) An utterance *u*, at time *t*, by speaker *s*, of "The hobbit over here isn't real" is true if and only if the object *o* indicated by *s* by her use of "over here", at time *t*, is (presents as) a hobbit, and isn't real.
- (iv\*\*) An utterance *u*, at time *t*, by speaker *s*, of "I am hungry" is true if and only if *s*, at time *t*, is hungry.<sup>22</sup>

<sup>20</sup> This isn't an ideal way of putting the matter. See Azzouni (2012a) for the best way to speak about this.

<sup>21</sup> It may be that the content of "she"—being female—must be demoted to "presents as female." This will depend on what properties are to be attributed to hallucinated objects; and this isn't something I can discuss further now. See Chapter 2 of Azzouni (2010b).

<sup>22</sup> (i\*\*)–(iv\*\*) are meant to be simple illustrations of a broadly characterized family of approaches. In particular, and among other things, in formulating them I'm not attending to issues about exactly how conditionalizations are supposed to be characterized, and in what detail; I'm skirting over issues about exactly how content that appears in the utterance (e.g., "vase", "he", etc.) is supposed to contribute to the truth conditions of the utterance, and I'm also skirting over mismatch issues between demonstrative non-referring expressions—not because an hallucinated object is demonstrated, but because (say) the object someone intended to demonstrate was moved elsewhere—and the definite description on the right side is false as a result. The data are complicated, and the options are numerous, debated, and debatable. See Lepore and Ludwig (2000, 230–238), for dis-

*Summary of This Section* It follows from the foregoing discussion of standard (and non-standard) truth-conditional semantic theories that although “truth conditions” do supply necessary and sufficient conditions for the truth of sentences, the actual truth-values of statements often cannot be read off of those truth conditions. Consider:

- (8) Clark Kent is depicted as being identical to Superman in almost all comic books about him.

This is true if and only if Clark Kent *is* depicted as being identical to Superman in almost all comic books about him. But the truth-inducing factors for (8) are to be found not among the facts about Clark Kent a.k.a. Superman (and how *he’s* depicted), because there is no such entity, but rather in the facts about our fictional (comic book) practices.<sup>23</sup> This means that the special science of semantics isn’t itself the whole story about language-world relations. Rather, it must be supplemented by a study of those relations themselves. And in the case of sentences with non-referring terms, those relations are intricate and complex, and so can’t be characterized by the simple laying out of truth conditions as it’s done in the tradition originally stemming from Tarski’s work.

Some may think that the science of semantics should encompass truth conditions in a substantially determinate sense, that the truth conditions that semantics provides for expressions should *really* characterize the possible circumstances of the truth and falsity of expressions in a way similar to how a description of a triangle as being an enclosed, three-sided figure composed of straight-line segments provides necessary and sufficient conditions for a figure being a triangle. They shouldn’t do so in the way that being a triangle provides necessary and sufficient conditions for being a triangle. I have additional reasons for doubting that truth-conditional semantics should be so conceived. But those reasons can be set aside for current purposes.<sup>24</sup>

I can, however, provide some illustrations of what I mean about the empirical boundaries of semantics stopping short of the actual assignment of truth values to sentences or utterances in a context. Consider the utterance “That’s Santa Claus,” when the child is pointing at a picture of Daniel Dennett, and contrast it with the same utterance when the child is instead pointing at a picture of Santa Claus. That a picture of Daniel Dennett (and thus Daniel Dennett) is within the (deferred) scope of the child’s pointing gesture makes *him* the truth-maker of her claim “That’s Santa

---

cussion and criticism of various approaches to the conditionalizations of complex demonstratives. It should be clear already, however, that my points about non-referring expressions in such truth-condition clauses will be unaffected by the replacement of my (i\*\*)–(iv\*\*) with appropriately complicated alternatives.

<sup>23</sup> Contrast this, for example, with the otherwise similar facts about Donald Trump and how *he’s* depicted (in various media). These are facts about *Donald Trump*—the object Donald Trump that *exists*. And notice this. In the case of real entities—like a Donald Trump—the reasons for *why* they’re depicted the way they are in such-and-such a place *may* have a lot to do with them. But that’s never the case for Superman, and other “fictional entities.” That Superman has such-and-such properties never explains anything. (There is no Superman and so he has no properties.)

<sup>24</sup> See Azzouni (2008), if you’re interested.



Claus,” in the first case, and the how-it-is with Daniel Dennett—that he’s not Santa Claus (as far as *I* know)—makes the statement false. But there is no truth-maker for the utterance, “That’s Santa Claus,” if a child points at a picture of Santa Claus. In this case, it’s a pictorial practice that induces truth values for this utterance, and that determines that it’s true.

In general, there is a lot of asystematic complexity in the relationships between what induces truth values in sentences and those sentences. To describe, for example, what induces a truth value in the child’s “That’s Santa Claus,” I only had to mention a certain pictorial practice (a picture-of practice that we all understand) with respect to a certain picture. Should the child utter, “That isn’t Mickey Mouse,” however, while pointing at a picture of Santa Claus, this will be a different matter. Here one must bring in an entire cartoon tradition, and describe how nonexistent things are depicted as alike and different; for it’s *those practices* that induce a truth value for the sentence “That isn’t Mickey Mouse” in this case. Something similar happens with “ $2 + 2 = 4$ ,” on the nominalist view. One has to invoke indispensability considerations about *entire areas* of applied mathematics. This may require the invocation of certain physical objects that mathematics is applied to as well as invoking certain systematic practices that we engage in.

The foregoing illustrates that even if a philosopher or linguist thinks it’s appropriate for a semantic theory to provide truth conditions, as it’s done in the Tarskian tradition, it doesn’t follow from this that it’s reasonable for a semantic theory to also provide a theory of how the resulting truths (and falsehoods) are due to the machinations of real objects in the world. This is because the relationship of our statements to the world only has a *hope* of being straightforward in cases where everything exists that the sentence is about.<sup>25</sup>

## 6.4 Tarski Biconditionals and Truth-Conditional Semantics—Applied to Crusoe 5’s Language

As I claimed in Sect. 6.2, Crusoe 5’s language is just like ours with respect to truth-conditional semantics. In fact, things are even *better* for his language than for ours because I’ve *stipulated* in the thought experiment about him that when his terms do refer (when the grounding facts *do* determine correspondence facts) none of the complexity of words like “London” arises, that Pietroski and Chomsky are so exercised about: the terms “coconut,” “yuckonut,” the counting words, and so on, have

---

<sup>25</sup> In my view, however, this hope is dashed with respect to natural languages (although not for certain artificial languages). See the discussion of Pietroski and Chomsky in Sect. 4.4, and recall the important claim, in particular, that it’s an empirical possibility that language-world relations are likely to be theoretically intractable—regardless of whether terms (as used) refer to real objects or not. I’ve borrowed much of the discussion in this section from Azzouni (2012b, 264), while modifying statements that were far too strongly put in that original article. (This last sentence is for readers interested in such matters.)

the simple referential relations to things and to collections of things so loved by truth-conditional semanticists. It's just that the grounding facts about Crusoe 5 always induce gappiness and other infelicities that shift over time (but pretty much never vanish).

Should Crusoe 5 be so inclined, therefore, he can avail himself of a truth-conditional semantics the way we can because his truth and falsity expressions are like ours. Indeed, presuming he has named two objects in his world, "Julie" and "Mike" and presuming he has the existential quantifier, "Something," as well as the ampersand, he can utilize the first (baby) truth-conditional semantic theory I gave in Sect. 6.3, here repeated,

"Julie is a coconut" is true iff Julie is a coconut.

"Mike is a coconut" is true iff Mike is a coconut.

"Something is a coconut" is true iff something is a coconut.

If  $S_1$  and  $S_2$  are sentences, then " $S_1$  &  $S_2$ " is true iff  $S_1$  is true and  $S_2$  is true.

What about Crusoe 5's use of the knowledge and ignorance expressions? It does seem, as the discussion in Sect. 6.3 indicated, that Crusoe 5, like us, can also help himself to them, and so he can impose (on his own language) a bivalent logic compatible with the use of those expressions. ("There is a specific number of coconuts in that tree—even though I don't know, and will never know, what it is.")

Here's a worry: Crusoe 5's way of talking seems to apply truth and falsity to sentences of a more radically indeterminate nature than to sentences with terms that don't refer. There are things about the *world*, after all, that are relevant to evaluating the truth values of sentences like (1), here repeated,

(1) Minnie Mouse has *never* been depicted in movies as a plumber.

It may even seem that one can operationalize a search procedure for determining the truth or falsity of sentences like this. But contrast this with a claim about the number of items in some collection that (by assumption) outstrips Crusoe 5's dispositional resources now and forever. Is it, nevertheless, supposed to be true even in this case that there is some number that is the number of items in that collection?

Yes, because this is like cases where sentences with non-referring terms aren't stipulated in their truth values by a practice. And in these cases, Crusoe 5 speaks (as we similarly do) of not 'knowing' whether something is the case or not.<sup>26</sup> Crusoe 5 takes it that every tree containing coconuts *has* a particular number of coconuts. This follows from his understanding that "coconut" refers to all and only the coconuts independently of his inclinations to so apply the word. He has the same understanding of the applications of cardinal-number words. Crusoe 5 doesn't recognize that the references of his words wax and wane according to the dictates of his best dispositions at a time; and so he has a complementary way of speaking of his ignorance, of his not knowing something, that makes no distinction between three kinds of cases: (i) where he doesn't 'know' something for which his dispositions (at a

<sup>26</sup>For the purposes of the rest of the discussion in this section, I'm now distinguishing Crusoe 5's words with c-quotes from otherwise homophonic uses of such words by God and by us.

time) *will* provide an answer if he investigates, (ii) where he doesn't 'know' something for which his dispositions (at some later indeterminate time in the future) will *then* be able to provide an answer, although none of his dispositions can do so now, and (iii) where he doesn't 'know' something although his dispositions can *never* provide an answer.

Crusoe 5's words, 'true' and 'false', it must be stressed, are *his* words (as the c-quotes graphically indicate). And so (according to God), just as the references of all his other words wax and wane to the dictates of his best dispositions, so do these words. Naturally corresponding to this practice with 'true' and 'false', as I've just noted, is a usage of ignorance, of not 'knowing', that doesn't distinguish between cases (i)–(iii).

Let's turn from issues about Crusoe 5's use of truth and falsity expressions, as well as his use of truth-conditional semantics, to questions about how God or we are to apply tools like these to Crusoe 5's language. Start with God. Should God accept the cogency of Crusoe 5's use of the words 'true' and 'false'? He can if He wants to, *by adopting Crusoe 5's language*. His understanding of Crusoe 5's language can take this form: 'It's true that there are such-and-such numbers of stars or there isn't.' However, although God knows of every sentence in His own language which is true and which is false, this isn't so of Crusoe 5's language. In particular, God doesn't know how to fill in the blank in the sentence of Crusoe 5's language which reads: 'the number of stars is ---.' God's omniscience is unlimited in His own language; but it's limited when He chooses to understand the words in Crusoe 5's language. His omniscience is limited because, in Crusoe 5's language, talk of what's true and what's false outstrips any grounding facts that determine the appropriate correspondences between sentences and the world.

I should point out that these claims have—strictly speaking—nothing to do with the fact that God is allowing Himself to understand Crusoe 5's *language*, as opposed to Crusoe 5's *thoughts*. The same is true even if Crusoe 5 doesn't have a language, but just thoughts and concepts. In this case, such thoughts and concepts are disposition-meaning thoughts and concepts, and so God recognizes that their scope and range goes just so far as Crusoe 5's dispositions allow them to. The absence of grounding facts for Crusoe 5's thoughts leads to the same result for God: When He chooses to understand Crusoe 5's thoughts, his omniscience is correspondingly limited: He doesn't know—with respect to many collections, 'how many there are.'<sup>27</sup>

The foregoing means that God has only three options for interpreting Crusoe 5's language or for applying a semantic theory to Crusoe 5's language from the vantage point of His own. One thing He can do, of course, as I just mentioned, is adopt (not translate) Crusoe 5's language and, correspondingly, a truth-conditional semantics for that language using the terms of Crusoe 5's language. The result, as I said, is a loss of omniscience: there are many sentences (of Crusoe 5's language) that God doesn't know the truth values of.

<sup>27</sup> In this case the c-quotes aren't signaling quotations of Crusoe 5's sentences but instead characterizations of his thoughts.

A second option is to stipulate an arbitrary translation and *impose* truth values on Crusoe 5's sentences based on how truth values for the translations of those sentences sort out in God's language. There are a specific number of coconuts in any tree in Crusoe 5's language relative to a translation into God's language. Of course, God's options for translation—because of the absence of grounding facts in Crusoe 5—are infinitely arbitrary. He has many choices, but no way to genuinely choose among them. This is just the rule-following paradox again. The grounding facts are missing, and so God can't legitimately stipulate non-arbitrary answers in lieu of those facts: He has no basis for any choice he makes.

A third option is to adopt an approach to semantics that I've argued for in other work. Give up on necessary *and* sufficient truth conditions, and instead give necessary conditions and sufficient conditions (based on Crusoe 5's dispositions at a time) that aren't necessary *and* sufficient conditions. This is a semantic theory that will be time-linked (or, a series of semantic theories indexed to time) because the truth conditions must change as Crusoe 5's dispositions change. This approach (unlike the standard Tarskian truth-conditional approaches to semantics) *won't* yield a translation of Crusoe 5's language to God's. Rather, it will allow God—up to a point—to describe semantic properties of Crusoe 5's language. I like this third approach best; but I can't provide more details for it now (or argue for its merits).<sup>28</sup>

Our approach to Crusoe 5's language is just like God's in our having the three options that He has. In practice, however, the third option may be intractable for us because we aren't omniscient and so we won't have the kind of knowledge of Crusoe 5's dispositions that God is guaranteed to have. On the other hand, if our dispositions match up more or less nicely with Crusoe 5's then we adopt the second option, engage in a straight translation of his terms to ours. The arbitrariness that God faces in choosing a translation isn't faced by us except insofar as we face such arbitrariness with respect to our own terms anyway—because of the original rule-following paradox.

I've spoken of God's loss of omniscience should He take the first option: choose to understand Crusoe 5's language. Of course, the same is true of us: God will suffer a similar loss of omniscience should he choose to understand our language.

## 6.5 Some Concluding Remarks

In this chapter, and in the preceding ones, I've shown that the private-language practices of an isolated individual (Crusoe 5, in particular, and most valuably) are compatible with a way of speaking of the truth and falsity of his sentences that—with respect to most of his sentences—outstrips the facts that can determine whether those sentences are true or false. This way of speaking is both available to him within his own language, as well as being available to others, outside of his language—if his language is translated into our language, for example. Furthermore,

---

<sup>28</sup> See Azzouni (2007, 2008).

his private language is compatible with a truth-conditional semantics, as I've shown in this section—both if applied by him (privately) or applied to his language by others.<sup>29</sup>

It's important to realize, in defense of my solution to the rule-following paradox, that *any* solution to the rule-following paradox that substitutes assertability conditions for truth conditions does more than just eliminate the cogency of private rule-following. For similar reasons, it also eliminates the cogency of important aspects of our self-ascribed ability to (privately) *think*. This is, of course, the main reason Kripke's Wittgensteinian sceptical-solution to the rule-following paradox is so shocking and dramatic. The result of it is that our thoughts can only be cogent in a setting where they correspond to sentences or other public vehicles of communication that we're entitled to assert (subject to the ratification by others). What's implied, therefore, is a kind of *public sentimentalism*: a thought is cogent only if it's normatively constrained by community standards. This corollary—I must stress—does *not* follow from my sceptical solution to the rule-following paradox.

Here's an important point I've tried to show in this book. As I mentioned in the first chapter, many philosophers have tried to draw a sociological conclusion from the rule-following paradox. On my reading, this mistake is a symptom of an overly-narrow appreciation of what's required of a coherent normative practice of correcting oneself and others—that a coherent practice of correcting “errors” requires a *fixed* standard that's external to the individual being corrected. This forces the view that such an external standard must be located in the community to which that individual belongs. And *this* motivates what I called, in Sect. 2.5, the straight sociological solution to the paradox. The other approach—Kripke's Wittgensteinian one—is more radical: it replaces truth with public assertability conditions. Because of the nature of such assertability conditions, practices of correction can only be cogently applied in a public setting. Thus a community standard is also required by that view.

When we think about our ordinary practices, however, we realize that there is a tension between them and *any* sociological solution—sceptical or otherwise. It's part of *our* practice to regard it as possible for *one person* to be right, and for everyone else to be wrong. We allow that the ability of one person to sort certain kinds of objects may turn on his having certain skills, and on his being able to detect differences in things that others aren't capable of seeing. We also allow that it's possible for it never to be recognized (by other people, or by the society at large) that such a person is “right.” More strikingly, we allow that it's possible for an entire society to

---

<sup>29</sup>I should note that these points—in a way—are anticipated by Kripke's Wittgenstein. Kripke (1982, 86) attributes to Wittgenstein the view that talk of truth and falsity, and even the use of a “calculus of truth functions,” is compatible with his sceptical solution to the rule-following paradox. That means, I must add, that this talk is compatible with a notion of truth *sans* correspondence or fact. The additional insight I attribute to the truth-deflationist is that a notion of truth *sans* correspondence or fact is actually all the “truth” anyone needs; that includes the need to use it in truth-conditional analyses of languages. (See Azzouni (forthcoming), for details.) Such a fully-functional notion of truth is only mistakenly seen as requiring either correspondence or facts.

be wrong, and for that society to never be corrected.<sup>30</sup> One way to imagine such a thing is to imagine a Friday whose counting practices enjoy a superior positive success curve compared to those of every other individual in that community, so that he's capable of exploiting anyone else in (for example) barter exchanges. (He can, that is, use the *entire society* as a money pump.)

My approach, despite being a sceptical solution to the rule-following paradox, can accommodate the possibility of either an entire society being wrong, or of an entire society being right, and one individual (who exploits the practices of that society) being right, because it bases the cogency of someone correcting her own practices in terms of her subsequent enjoyment of an enhanced positive success curve in her interactions with the world. The cogency of someone being right or wrong, therefore, is based in the interaction of her developing dispositions with her environment, *regardless of whether she realizes this or not*.

It's worth reminding ourselves of the powerful intuition we all have that a Robinson Crusoe is capable of developing a coherent language (and accompanying concepts) that he can apply successfully to the objects in his world in conformity with his previous intentions, and yet is one that's not to be understood as relative to the standards of any community. As I've already stressed, it's these intuitions that make the "private language argument" so shocking in its import. Indeed, we all have the powerful intuition that our (own) grasp of counting enables us to understand how to go on *ad infinitum*.<sup>31</sup>

My approach, being a sceptical one, ratifies these intuitions only in the very weak sense that it's *true* that we (and Crusoe) have the experience of "understanding" how to go on. There is no matter of fact about this "understanding"—about our capacities—that fully underwrites this truth. What this shows, in turn, is that our ordinary talk of "understanding" (since it also belongs to our language) is just like all the other words in our language in going beyond "the facts," what, in this case, any study of us—psychologically and subpersonally—can ever reveal. "Understanding," is a perfectly ordinary word (not an especially philosophical word) that we use to indicate confident competence at certain tasks and in certain

---

<sup>30</sup> Kripke describes his own misgivings about Wittgenstein's sceptical solution this way: "But may the individual doubt whether the community may not in fact *always* be wrong, even though it never corrects its error? It is hard to formulate such a doubt within Wittgenstein's framework, since it looks like a question whether, as a matter of 'fact', we might always be wrong; and there is no such fact" (Kripke (1982, 146), italics his). Kripke then notes that he has avoided a more extensive discussion because in doing so he "might have to abandon the role of advocate and expositor in favor of that of critic."

<sup>31</sup> Kripke (1982, 21–22) writes, "Sometimes when I have contemplated the situation, I have had something of an eerie feeling. Even now as I write, I feel confident that there is something in my mind—the meaning I attach to the 'plus' sign—that *instructs* me what I ought to do in all future cases. I do not *predict* what I *will* do ... but instruct myself what I ought to do to conform to the meaning. (Were I now to make a prediction of my future behavior, it would have substantive content only because it already makes sense, in terms of the instructions I give myself, to ask whether my intentions will be conformed to or not.) But when I concentrate on what is now in my mind, what instructions can be found there?"

ways. It should be no surprise, therefore, that like “coconut,” like the number words, and like pretty much all our words, the standards for it also *wax and wane*.<sup>32</sup>

In the case of “understanding,” the way the word waxes and wanes with respect to what it applies to at a time involves (in the case of counting, for example) later, and better positive-success-curve-inducing, dispositions due to people learning first how to count with their fingers, and then to count by means of various tools (calculating devices: counting boards, etc.; notational devices: Arabic notation, etc.)—eventuating in the use of sophisticated devices of all sorts. At each stage in our acquisition of new approaches to counting, we gradually iron out various “mistakes”: various earlier methods of counting that induce a less than optimal positive success curve. We don’t think of what’s happening *this way*, of course. We think instead that we are continuing to count in the same way we always intended to count. And (factlessly), *we are*. I will develop the contrast between these two ways of thinking about our language (or Crusoe 5 thinking about his) in the next chapter.

---

<sup>32</sup>I revisit this theme of waxing and waning with respect to what might be called “higher-level” terms in the next chapter.



## Chapter 7

# Correspondence Metaphysics and the Cogency of a God's Eye View

**Abstract** This chapter addresses two remaining issues. The first is the apprehension that in undercutting reference-magnetism views the way I have, I've ruled out the possibility that the way the world sorts out for us empirically could be compatible with some version of correspondence metaphysics—an empirically-justified neat correspondence between language (kind terms, in particular) and the world. Previous chapters may also have given rise to the worry that my approach methodologically presupposes a God's eye view while at the same time undercutting itself by officially regarding that viewpoint as one that neither we nor private rule-followers can make sense of. These issues are connected and I explore them both in this chapter. I also characterize two powerful and apparently conflicting ways of thinking about language, the contrastivist view and the noncontrastivist view. I don't try to resolve which view is the right one.

### 7.1 Introduction

My sceptical solution to Kripke's (Wittgenstein's) rule-following paradox, as well as many of its implications, should be clear. The key elements have been spelled out in previous chapters. First, the need for plpci dispositions in a private rule-follower to make possible a coherent private mistake-and-correction practice has been shown. Such a coherent private mistake-and-correction practice makes possible self-ascriptions of truth and falsity, as well as a self-applied truth-conditional semantics. I've also indicated (in Chap. 6) to what extent outsiders can use the same apparatus—truth-conditional semantics—on the private rule-follower's language.

This chapter, therefore, is dedicated to the sundry tasks of mopping up a couple of loose ends—some concerns that (I think) some philosophers may still worry I haven't addressed, as well as some items that have been raised on several occasions when I've discussed my sceptical solution to the rule-following paradox.

One important issue I'll address is the apprehension that in undercutting reference-magnetism views the way I have, I've ruled out the possibility that the way the world sorts out for us empirically could be compatible with some version of correspondence metaphysics—an empirically-justified neat correspondence between language (kind terms, in particular) and the world. I hope the previous

chapters of this book—especially Chap. 4—don't give this melodramatic impression. Even so, I want to say a little more about this in concluding the book.

It has also been suggested to me that there are reasons to think that my approach methodologically presupposes a God's eye view while at the same time undercutting itself by officially regarding that view as incoherent—or at least officially taking it to be a viewpoint that neither we nor private rule-followers can make sense of. (And what a philosopher can't make sense of isn't something she's allowed to presuppose.) These issues are connected. I explore this second issue in the next two sections, and then turn to the first one.

## 7.2 Contrastivist and Noncontrastivist Views of Private Languages

Here's one worry some philosophers may raise.<sup>1</sup> I've apparently introduced a standard of objectivity via a God's eye view of how our words should be applied that contrasts with how we actually apply our words. I've also suggested—in previous chapters—that Crusoe 2, Crusoe 3, and Friday don't have access to God's perspective. More dramatically, they apparently can't even make sense of certain crucial notions that seem needed to characterize a God's eye view: they can't make sense of words better fitting (or worse fitting) the world. They can, of course, contrast one language with another by the varying positive success curves they induce; and they can fault the dictates of one language in terms of another by translating the terms of one language to that other, and treating the resulting deviations in truth values as mistakes.

This point can be put—and it has been put by many philosophers—this way: No one, including us, can escape their own language frameworks, conceptual schemes, or whatever. But when the point *is* put this way, then it seems to imply that Crusoe 5 (and ourselves) also can't make sense of the idea of a language fitting better or fitting worse to the world and (correspondingly, therefore) neither Crusoe 5 nor ourselves can make sense of a God's eye view of the world either. Thus—so the objection goes—a God's eye view has been shown to be incoherent. Any contrast between such a perspective and our own is, as a consequence, nullified since we can't make sense of God's end of the contrast. Therefore, the best sense of anyone's words being right or wrong can only be in terms of how we use our words in our (current) language framework, and our ability (which the Friday of Chap. 2 shares) of faulting—relative to a translation—the words of others when they deviate from our own words. I think, however, that the matter is more complicated than these considerations make it sound. Although I've indicated my views about this in Sect. 3.4, the point deserves a full discussion.

I'll start with the point that our picture of how the words of our own languages operate is largely the same as Crusoe 5's. Let's consider the coconut again. Our term

---

<sup>1</sup> I owe much of particularities of how I pose this worry to Douglas Patterson—9/24/09 email.

“coconut” refers to coconuts. We take this to be the case even though we treat our dispositions—even our collective dispositions—as all potentially untrustworthy. This, furthermore, is how we understand the relationship between (nearly) all our words and our dispositions.<sup>2</sup>

One view—I’ll call it *the contrastivist view*—takes it that this book has shown that this picture is just *wrong*. The right view is that we (and Crusoe 5) utilize a continuously-changing series of disposition-meaning idiolects, where each of the words in each of these idiolects refers exactly to what the dispositions of the speakers of those idiolects (at those times) impel it to refer to. We (and Crusoe 5) suffer from the ineliminable illusion (because our dispositions to apply words are largely subpersonal) that the words in our language are largely stable in what they refer to.<sup>3</sup> We—but not Crusoe 5—think of our language as *public*<sup>4</sup> although we and Crusoe 5 agree that words are frozen in what they refer to just to the extent that they are fixed in their meanings. Due to great variations in knowledge (both among shared-language speakers, and over time for any particular language speaker), however, we differ greatly in our abilities to apply those words correctly. “Coconut” is an example of a word that—we think—doesn’t change in what it refers to over time; rather, we learn (over time) that some of the things that we thought were coconuts aren’t in fact coconuts (and vice versa). According to the contrastivist, the fact that our words and their references wax and wane over time is one we misperceive (and misdescribe) as an epistemic fact. We think, instead and incorrectly, that our *knowledge* of which objects our words refer to and don’t refer to is what waxes (or wanes) over time and that the references of our words don’t do this.

Corresponding to these misconceptions is a semantic theory that we misguidedly apply to our language. Such a semantic theory, for example, presupposes that a certain class of declarative sentences is two-valued,<sup>5</sup> and it assigns truth conditions to those sentences along standard Tarskian lines, based on what the words in those sentences refer to.

According to the contrastivist, the semantic reality is very different. Each person’s plpci dispositions induce disposition-meanings (for that person) for each sentence S: such-and-such circumstances when each individual would assert S, deny S, or fail to respond in either way (e.g., by an “I don’t know” shrug, or by some other failure to assert or deny S). These disposition-meanings change over time because of the posi-

---

<sup>2</sup>The ways we, in the first-person, apply certain words—“pain,” etc.—strike us as exempt from this. I provide a naturalistic explanation of this later in this chapter.

<sup>3</sup>We *are* aware, of course, that words can change in what they refer to (they can change their “meanings”). But we see this process as sporadic and somewhat slow (e.g., generational)—the kind of process, for example, that William Safire notoriously (and weirdly) denounced.

<sup>4</sup>That is, an illusion that we have, but that Crusoe 5 doesn’t have, is that language is a public object; we don’t experience the reality that each of us speaks our own idiolect (which is at best only similar to the idiolects of those around us who we talk to). See Azzouni (2013).

<sup>5</sup>I’m here abstracting away from the rich complexity of natural languages, and—of course—from the more complex semantic theories that are available to handle that complexity. In the interests of keeping things simple, we might, for example, restrict our attention to simple statements of the cardinality of collections of objects.

tive way that these changes enable individuals to succeed in their environments, and to communicate with one another. The nature of such changes is what (i) causes those individuals to enjoy the illusions of a shared stable language I described in the previous two paragraphs, and (ii) allows speakers of these changing idiolects to instead speak both of the truth and falsity of sentences and of the truth conditions for such sentences. In reality, however, sentences have no truth conditions, at least not in the way that misleading Tarskian semantic theories take there to be. Instead, truth and falsity conditions (this is one possibility) for each idiolect arise from the plpci dispositions of a speaker of that idiolect at the time that speaker possesses that idiolect.<sup>6</sup>

It might be thought that I've too narrowly described the resources that plpci dispositions offer for underwriting semantic theories. We can take a chronologically longer view of those dispositions and characterize truth conditions only by the standards of later dispositions. Only with respect to the later dispositions, and only when they eventually stabilize on verdicts for certain sentences, do we take those verdicts to indicate the actual truth values of those sentences. Doing this, however, still leaves numerous sentences without truth values; furthermore, we don't normally think of a stable verdict on a sentence as necessarily a correct verdict.<sup>7</sup>

Lastly, the contrastivist diagnoses a certain mistake about the standards for our words that arises from the foregoing false picture of our words: that the world provides the normative standards for the applications of our words. "Coconut," for example, is correctly applied to something if and only if it *is* a coconut. The real story of where the standards for our words are coming from has to be told in terms of the dispositions of the private rule followers themselves, and how those dispositions evolve in response to the positive or negative impact of the world on those private rule followers.

The contrastivist view that I've been describing in the past few paragraphs posits a sharp contrast between the reality of our languages and their relationships to our dispositions and the illusory picture of these relationships by speakers who deeply misconceive them. The contrastivist view is one I've repeatedly attributed to God—but that doesn't make it *right*. There is another possibility, what I'll call the *noncontrastivist view*. The noncontrastivist denies that God's perspective of our languages falsifies our perspective of them. The relationship, rather, is like that between faces as they look to us normally, and what faces would look like if one instead saw the anatomy beneath the skin. The perspective in terms of plpci dispositions (the perspective from the point of view of "the anatomy of reference") describes what might be called the "engineering" of reference. But, despite this enhanced perspective, our words nevertheless refer to what they refer to, and part of the explanation of how they manage this involves the trajectory over time of our (collective) plpci dis-

<sup>6</sup>Notice that "truth conditions," as the contrastivist insists they be understood, don't correspond to how the speakers use their words "true" and "false." Speakers do *not* treat their assertions at a time as true by virtue of their dispositions to so assert them. This is because—on the contrastivist view—they're presupposing a false picture of their language(s).

<sup>7</sup>One thing we can't do is characterize the truth conditions of sentences in terms of dispositions of people to assert or deny sentences when a "mistake" isn't being made. This would only be to again face the rule-following paradox at the level of semantic theory: the notion of a "mistake" can only have content in relation to the dispositions of speakers at some other time.

positions. The noncontrastivist view takes “coconut” to refer to coconuts. This is a reference relation between a kind term and objects in the world, one that’s named by the words “refer,” “reference,” and so on. The “correspondence relation” is the relationship between the word “coconut” and certain objects in the world—coconuts.

Recall the distinction between grounding facts and correspondence facts that was given in Chap. 1. The noncontrastivist view is that the correspondence relation is only *partially* supported by the grounding facts, and as a result it’s only *partially* instantiated by correspondence facts. That is, the absence of (certain) correspondence facts arises from absences in the grounding facts in individuals (taken individually or collectively). For, first, as we’ve seen, the (individual or collective) dispositions to understand how to apply one’s words in certain contexts may simply determine nothing at all about a great number of things—for example, whether they are coconuts or not. Second, the dispositions may determine falsely (from a later perspective) that items are (or aren’t) coconuts.

In describing the reference relation this way, the noncontrastivist simply extends to the very word “reference” itself—at least as *we* use it—how we (and Crusoe 5) talk about the numbers of coconuts in trees, and other similar claims about what is or isn’t so-and-so (independently of whether we can ourselves determine it or not). Therefore (according to the noncontrastivist) it doesn’t have to be denied either that “coconut” refers to coconuts, or that “refers to” characterizes the reference relation. Rather, the noncontrastivist is claiming that only *some* instances of the reference relation go beyond the facts.

To make the noncontrastivist view clearer, it’s valuable to distinguish the *applications* of a word from what it *refers* to. What a word applies to (at a time) is what the (collective) dispositions of its users (can) apply it to. On the noncontrastivist view, what *waxes and wanes* isn’t the reference of words, and it isn’t (*pace* the contrastivist view) that words in languages are constantly superseded by successor words in successor languages. Rather, it’s the applications of words that wax and wane. Their references remain stable (unless the word is discarded or changed in what it refers to).

On the noncontrastivist view, reference does *not* wax or wane; no more is it that words are continually replaced by new words with differing extensions. (The noncontrastivist doesn’t deny that meaning change occurs; she thinks, however, that there is a sharp difference between changes in the *application* of a word and changes in its meaning.) But the standards by which these words are applied and, correspondingly, how mistakes are recognized isn’t to be characterized in relation to what such words refer to but to the waxing and waning applications of these words. Therefore, just as on God’s view, the *standards* by which our words are applied *wax and wane*. God has it right about how we apply our words and how we recognize ourselves to have made mistakes. But, the noncontrastivist claims, *pace* God, that we shouldn’t conceptually connect our practice in implementing those standards with what the words refer to—in the sense that what the words *refer* to is to be the standard by which our words are to be judged as correctly or incorrectly applied.

As I suggested in Sect. 5.4, this doesn’t mean that it would be right to say that there are *no* external standards for the use of our words. The suggestion would be that reference-magnetism views fail precisely because they try to provide *local*

word-to-world characterizations of the external standards for words: it's specifically the coconuts (or the natural kind of coconuts) that are what's supplying the standards for our use of "coconut." This can't be right. If our word "coconut" after waxing and waning for a while settles on a permanently-stable collection of objects to refer to, that's a matter of empirical luck enabling a nice matching of our dispositions to apply a word to a collection of things in the world. This nice matching can't be treated, however, as playing a causal role (or any role, for that matter) in how the standards for the references of our words are established. Relatedly, that nice matching can't be the explanation (or ground) of the standards that we're using to apply the word. Those standards would have been in place even if this wonderful final state of affairs didn't emerge (because the world unkindly precluded it, say). In any case, the world intrudes, and provides standards, *only* in the sense that it globally rewards (or globally punishes us) for the results of our dispositions, and how those dispositions evolve over time. And, correspondingly, things only work out for us (in time) if we have the neurophysiological capacity to induce plpci dispositions in ourselves (that induce positive success curves in our interaction with the world).

On the noncontrastivist view, we understand the application of a word to be its reference when there are no longer any mistakes being made. But "mistake," as here used, is similar to "not knowing," as characterized earlier: the words go beyond any description—now or ever—of how someone could modify their dispositions to yield a different (i.e., "correct") answer. (*All* our words, on the noncontrastivist view, are at best *partially* supported by the correspondence facts in what they refer to.) This *negative* characterization of the reference of a word, therefore, allows us to speak—for example—of there being a certain number of items (regardless of whether there are any grounding facts whatsoever that underwrite what that number is). Similarly, we can speak of words *referring* to collections of objects regardless of whether there are any grounding facts that (fully) underwrite that reference relation.

In general, there are various possibilities. The (collective) dispositions to apply a word may wax and wane for a time before the word largely stabilizes in how it's applied. Our dispositions to apply that word no longer change although—of course—that isn't true of our dispositions to apply other words. Furthermore, defeasibly, these dispositions actually characterize what the word refers to. Words like "pain" are ones that in principle can be quite stable *without* having gone through a waxing and waning process—the states in question and dispositions to use certain words (or have certain concepts) are coordinated, nearly enough, by evolution. This is because there can actually be a tight causal relation between certain states of the subject and their disposition to apply the word "pain." This kind of locality restriction between states of the subject (that they all occur in the subject) holds generally of introspective terms; and can explain the stability of such terms, in contrast to words that are applied externally by a subject. This doesn't imply, of course, that dispositions to apply these introspectively-stable words can't be destabilized—neurological disorders (among other possible ailments) as well as artificially-induced experiences (like out-of-body experiences) show otherwise.

Another case is where a word stabilizes in how it's applied with respect to various items, and as our dispositions improve, it remains frozen on those items but

continues (forever) to accrue new ones. Our ability to apply ever-larger cardinal numbers to ever-larger collections of objects fits this case (with respect to the general phrase, “number of”). Again, this growing set of dispositions, defeasibly, actually characterizes larger and larger collections of items that cardinal words can be applied to. Lastly, there can be words that never stabilize in their application. Certain terms of failed scientific theories can look like this, I suspect, perhaps also terms like “ghost” or “witch.”

If a word can be applied usefully at all, in general, this is because our dispositions to apply that word are stable on at least some class of objects, and it may become stable on larger classes of objects over time. There are always appropriate areas of the application of words, however, that elude stability (and even application), and that will always do so. Nevertheless, just as we speak of the possibility of something being a coconut, or not—regardless of whether our dispositions (now or ever) to use the word determine an answer—so too, we speak of words referring to objects independently of whether there are any grounding facts, now or ever, that fully underwrite the correspondence facts that in turn fully determine that reference relation.

In Chap. 6 I’ve discussed what I call the “truth maker assumption”—the view that if a sentence is true, then it’s true because there are things (that the sentence is about) that it’s true of. There are many different versions of this assumption but they share the perspective that a sentence that isn’t about anything can’t be true or false. If a sentence *is* true (or false), then there is something that it’s true (or false) *of*. This assumption is one that I think we should resist; and the noncontrastivist view reveals a way of resisting it that’s different from arguments against it that I’ve given in other work. There are no correspondence facts about whether “coconut” refers or not, with respect to many items in the world—for the grounding facts about how we use “coconut” don’t determine such facts. *Nevertheless*, it’s true (or false) of each such item whether or not “coconut” refers to it.

This is simply because “refer”—and all our semantic notions, for that matter—are part of *our* language (and Crusoe 5’s corresponding notions, of course, are part of his language). Thus bivalence *applies* to these terms as well; and this is also true of the knowledge/ignorance idioms I discussed in Chap. 6.

I should stress the point that the noncontrastivist view isn’t restricted to sentences and language: it extends to our view of our own thoughts and self-ascribed concepts. We understand our concept of *coconut* to pick out all and only the coconuts—even if we *can’t* (and even if there are no facts that determine what all and only the coconuts *are*). Built into our concepts, into our understanding of our concepts, that is, is that our concepts apply to what they apply to even though the grounding facts about us don’t determine appropriate correspondence conditions (appropriate necessary and sufficient conditions) to sort everything in the world into what the concept applies to and what the concept doesn’t apply to. In the more ordinary language of Chap. 2, we take our concepts to be ones we can grasp even if (as in the case of calculation) we’re not particularly good at successfully applying them.

The last point I’ll make (to round out this section) is that the Tarskian approach to “true” and “refers,” as well as truth-conditional semantics, fits very nicely with the noncontrastivist view. This is because the deflationist notions of “true” and “refers”



function successfully regardless of whether the terms (e.g., “Mickey Mouse”) they apply to refer or not, and regardless—if the terms do refer—of *how* they refer. Because “true” and “refers” on truth-conditional approaches to semantics aren’t sensitive to the presence or absence or nature of language-world relations, they aren’t sensitive to the ways in which grounding facts can fall short of providing a basis for word usage.

### 7.3 Is Our Concept of a God's Eye View Coherent?

Recall the concern that the last section opened with. I gave the impression in previous chapters that Crusoe 2, Crusoe 3, and Friday don’t have access to God’s perspective. More dramatically, Crusoe 2, Crusoe 3 and Friday can’t make sense of certain crucial notions that are needed to characterize a God’s eye view. Specifically, they can’t make sense of words better *fitting* (or worse *fitting*) the world.

The noncontrastivist view creates conceptual room for the idea that our words refer to what they refer to *despite* our epistemically-faulty dispositions. In exactly the same way, conceptual room has been created for the idea of a Being whose dispositions aren’t epistemically faulty the way ours are: a Being who simply knows enough so that He can apply His terms as they refer.<sup>8</sup> In so imagining His words suitably applying to the world, it isn’t that we are to imagine the world itself as shaped such and so (as it were) and His words as conforming to those shapes. To do this really would be to step outside our own conceptualizations and to try to compare His conceptualizations to the world. So too, if we did something similar with ourselves—imagined our words as fitting well or ill with the world’s own contours—we would be illicitly stepping outside our own conceptualizations and trying to imagine (from outside those conceptualizations) a comparison of our own concepts with the world. What we *can* imagine is that the waxing and waning processes—the changing applications to the world that our words go through as our dispositions are refined—can come to an end. And we can imagine that they have come to an end with respect to a word because the positive success curve induced by our current dispositions is already maximal; nothing will make it better. Similarly, we can imagine that God’s conceptualizations are already maximal in the positive success curves they induce with respect to *all* His words (and concepts). Not even *He* can do anything to make them better.

Characterizing our notion of our words “suitably applying to the world” in terms of the concept *maximal success* doesn’t involve an implicit picture of a world that we match our words to, nor does it allow the illicit thought that we can use this so-called *suitably applying to the world* as the standard by which we can take our words to refer.<sup>9</sup> The standard for correcting our referential mistakes is always the same one:

---

<sup>8</sup>A Being, perhaps, whose experience of coconuts—and everything in the world, for that matter—is like the human experience of *pain*.

<sup>9</sup>So it’s probably wise to *drop* phrases like “carving the world at its joints” or “fitting the world as it is” because they are treacherously misleading for metaphysical thinking. The vanilla “suitably applying to the world” glossed—as I suggest—in terms of maximally-positive success curves is better.

introducing a more positive success curve; and we always treat our applications of our words at any time as defeasible: they may not really refer to what we think they refer to because the positive success curve in question may not be maximal. We can make sense of a Being who need not use His words defeasibly, nevertheless, by understanding His practices with His concepts as already maximally successful.

Recall the discussion in Chap. 3, where the question was raised whether Crusoe 2 can make sense of a language better fitting the world. One theme—stressed there and throughout this book—is that such a metaphysical notion can't function as a standard for how words are used. The noncontrastivist accepts this. In claiming that we can speak of what our words refer to—apart from how they are applied—it isn't therefore being claimed that what they refer to can be used as a standard of correction for our words. The standard for the correct usage of our words is one that waxes and wanes (it's one that the world induces by *globally* affecting our well-being). The second important theme raised in Chap. 3 is that it isn't possible for Crusoe 2 (and *a fortiori*, for any of us) to think of a language as “fitting” the world. I've pressed, instead, the idea that our notion of God's application of His words needn't be understood in this metaphysically loaded way. Instead, as noted, we can think of Him as epistemically faultless: as using his concepts (and words) in such a way that His positive success curve is maximal. We can accept the fact that there is no further content to our notion of words—ours or His or anyone else's—“fitting the world” that goes beyond epistemic faultlessness: that the divisions the words in question introduce induce maximally-positive success curves.<sup>10</sup> The worry was raised at the beginning of this section that the best sense anyone (us or Friday, for example) can make of someone being wrong is by comparing their usages of words to our own—relative to a translation—and criticizing theirs for deviations. But we can now see that this is wrong. Thinking of “objectivity” as an epistemically faultless God's eye view of things suffices to provide us with a notion that goes *beyond* our own current usage and dispositions to apply words. Thus this characterization of “objectivity” doesn't amount to the empty ratification of how we currently apply our *own* words.

What has been shown in this section is that the question of whether or not we can make a kind of sense of the idea of an epistemically faultless application of our words—say by an epistemically faultless Being such as God—comes down to whether the contrastivist or noncontrastivist view is the right one. On the contrastivist view, the idea of such a Being arises from adherence to the wrong picture of our language. The right picture of our “language” is that we actually shift through a series of disposition-meaning languages over time, and not that we have one language, that we apply the words of by our waxing and waning dispositions, along with waxing and waning standards of the correct application of those words. The noncontrastivist view opposes this by demoting the time-relative disposition-meanings to mere application-conditions of words at a time. The references of our words are instead characterized as those usages of them, if any, that would induce maximally-positive success curves.

<sup>10</sup> If someone—counter to my suggestion in footnote 9—insists on continuing to use the metaphor “fits the world,” *this* is how it must be understood.

What if there are no maximally-positive success curves (what if every positive success curve can be improved on)? Here are two empirical possibilities. The first is that all the successively more positive success curves induce application-conditions that agree on some of the applications of some of our words. Then, defeasibly, those words refer to those agreed-upon items. The second possibility is that, for each item that a given application-condition applies a word to, there is a later (more positive) success curve, after which the application-conditions induced by later, and superior, success curves exclude it. Then the still-valuable talk of *reference* is underwritten (ultimately) by *no* metaphysical correspondence facts. This is an empirical possibility, I stress: there is no way to *a priori* exclude it by philosophical reasoning of *any* sort—talk of the needed explanations of our success won't do it, because such success is always provisional. (I'll again stress this aspect of the human condition in the concluding paragraphs of this book.)

The important point being made here is that the noncontrastivist doesn't care. If this last case emerges as what's going on, the noncontrastivist says, simply demote the characterization of a "partially grounded reference relation" to the empirically accurate "ungrounded reference relation."

As this and the last section make clear, the contrastivist and noncontrastivist views don't look like terminologically disguised versions of one another.<sup>11</sup> According to the contrastivist, we think we speak one language, but we speak a series of them over time: we are in the grip of an illusion about our language.<sup>12</sup> According to the noncontrastivist, we speak one language just as we think we do. Furthermore, reference is "stable," just as we think it is. If we are confused, the noncontrastivist claims, it's about the nature of the reference relation: we mistakenly think that it's always determined by facts when this is at best only partially the case (as with *all* words)—subject to the possibility of the null case of complete ungroundedness that I raised in the last paragraph.

Now consider the philosopher who wants to claim that the notion of a God's eye view of the world (a view of the world, that is, that isn't epistemically faulty) is incoherent, and that even the gloss of it offered by the noncontrastivist is incoherent. It seems to me that it has been shown that this incoherence claim faces three obstacles. The first obstacle is simply that the noncontrastivist view must be shown to be *wrong*, that is, it must be shown that what waxes and wanes must be the *references* of our words and not merely the dispositions to apply those words—that the *reality* is that we speak disposition-meaning languages. This opposes the noncontrastivist picture that the reference relation (like the relations of *all* our words to the world) is at least par-

<sup>11</sup> But if they are terminologically equivalent, there's no harm in that. Speak as a noncontrastivist does rather than as a contrastivist does: it's more soothing. (I've been echoing pragmatist sentiments, especially in my invocation of *plpci* dispositions—and apart from what I say about the concept of truth. Now I'm apparently echoing William James because I'm not deciding between the correctness of the contrastivist and noncontrastivist, but instead offering a choice between them based on psychological comfort.)

<sup>12</sup> And that's no disaster, because we labor under *lots* of illusions about the properties of our language—many of them facilitate our successful use of our language. See Azzouni (2013) for an extensive discussion of these various involuntarily experienced illusions, such as the projection of meaning properties onto semantically inert designs. The psychological facts about these illusions are independent of the question whether contrastivists or noncontrastivists are right.

tially factless; and this, I want to suggest, is harder than it looks. It's hard to see, that is, what can be brought to force the interpretation of "reference" against the noncontrastivist—especially since the term "reference" is part of our (or Crusoe 5's) language.

Given, however, that the noncontrastivist *is* wrong, two obstacles still remain. The first is this: to show that the idea of a God's eye view is *incoherent* (because it arises out of the wrong picture of our language) requires showing that in some strong conceptual sense a view of language that allows the possibility of a God's eye view is *inconsistent*. It's not enough to show that it's been empirically ruled out. I'm not at all sure that a God's eye view *must* be taken to be "conceptually analyzed" in a way that requires it to have substantive content that goes beyond the negative characterization of it in terms of optimal success curves. But whether this is so or not has to be sorted out *before* the incoherence charge can be entertained, let alone sustained.<sup>13</sup>

The second obstacle is this. Suppose the perspective on language that the contrastivist treats as an illusion—that reference is stable and that our shifting relationship to it is an epistemic one—is one that it's psychologically impossible to shed. I think that the discussion of Crusoe 5 indicates why this would be true of *us*: the largely subpersonal operation of our, and Crusoe 5's, meaning urges prevents either him or us perceiving the references of our words to be constituted by our individual or collective dispositions. And suppose, for similar reasons, that it's also impossible to shed the same impression about our concepts. It then follows—because of this illusion—that it's built into our *conception of* our concepts that there is an in-principle right way that they can be applied. One way to see this neatly is to notice how easy and automatic it is for us to *understand* that how we are applying our concepts can almost always turn out to be wrong. It's not clear how we are supposed to be able to *understand* this, given that the contrastivist view rules out an objective (epistemically faultless God's eye) view—even one that's to be construed negatively in terms of optimal success curves (ones that can't be improved upon).

For given that we *do* understand our own application of our words and concepts to be fallible (apart from "pain," and the like), it follows that we also understand how—lacking that fallibility—we would have a God's eye view of the application of these words and concepts. But then, how can it be that a God's eye view is a notion that we *don't* understand (at least in the way that the noncontrastivist offers an understanding of it)?

## 7.4 Epistemic Modesty

We're currently (but eternally) in the position of not knowing what the future will bring. Even if progress in science has come to an end, as far as we can see—even if that science seems to be "final science," maximally optimal in its laws (wonderfully unified, maximally applicable—all that good stuff), we still won't know if our categories that are used in that science (and in the explanations we give on the basis of

---

<sup>13</sup> There is a cute analogy between this issue and what's called "negative theology."

that science) really do carve the world as it is in itself. (This is presuming, for the moment, that we can make sense of this metaphor—"carve the world as it is in itself.") This is because we never know *for sure* whether such a science—despite appearances—is a local product due to our being trapped in a local area of the universe that makes all the laws we have special cases.<sup>14</sup> There is no way out of this epistemic restriction because the real content of "carving the world as it is" can only be: our laws seem to be optimal *at the moment*. That this could change in a flash was noticed by Hume a long time ago. He's celebrated (to the extent that philosophers ever get to be celebrated) for this insight. Why should *anyone* (even a philosopher) think we can do better than provisional justifications of this sort? Why think, in particular, that by aprioristic talk about "explanation" or by trying to force such inductively established (epistemically frail) sets of categories to do extra philosophical work (in semantics!), we'll move a single inch away from where we're trapped: having induction as our only epistemic tool to establish laws of any sort, and to establish the corresponding categories that those laws are couched in?

Here is the final thought of this book. Hume's problem of induction is pretty old. Kripke's Wittgensteinian rule-following paradox, on the other hand, is fairly new. If I'm right—if, that is, my sceptical solution to the rule-following paradox is right—then the rule-following paradox emerges directly from Hume's problem when we realize that the human condition that Hume unearthed, about the epistemic frailty of inductively-established generalizations, *generalizes*. It generalizes to the very categories and terms that we use to express those generalizations, and that we apply to the world, and it generalizes (of course) to any abilities we have to facilitate those applications. *Any* knowledge that we think we have—even of language itself (even of mathematics, by the way)—is provisional in *so many* ways. We have, that is, *so many* reasons to be insecure and nervous (and frightened) about the future.

(But why end a book on such a depressing note? So let me try for something more cheerful.)

I'll conclude the book by concretely illustrating one way that the Humean insight about induction directly influences an aspect of our experience of rule following. I noted back in Sect. 5.3 that, because the dispositions (the mechanisms) that underwrite our grasp of learnt rules are subpersonal, we experience a sensation of being guided by those rules instead of experiencing them—to one degree or another—as simple urges. But more can be said about this. Our sensation that we "know how to go on" takes the particular form it takes precisely because these dispositions that enable us to competently count are inaccessible to consciousness. The inaccessibility of these dispositions means that we have no interior (introspective) grasp of the scope of our capacity to count. Consequently we engage in "premature generalization." Or, this is a better way to put it, we experience the results of an induction directed towards our own capacity to count as instead the grasping of a concept that we intend to execute. The result is an experience of unwarranted confidence about the future: we know how to go on. Sometimes we do, of course. But sometimes we don't.

---

<sup>14</sup> I think we can't even treat this locality possibility with respect to entire science as one with "low probability"; but that's not something I'm going to argue for here.

# References

- Azzouni, Jody. 1994. *Metaphysical myths, mathematical practice: The epistemology and ontology of the exact sciences*. Cambridge: Cambridge University Press.
- . 2000. *Knowledge and reference in empirical science*. London: Routledge.
- . 2003. Individuation, causal relations, and Quine. In *Meaning*, ed. Mark Richard, 197–219. Malden: Blackwell.
- . 2004a. Theory, observation and scientific realism. *British Journal for the Philosophy of Science* 55: 371–392.
- . 2004b. *Deflating existential consequence: A case for nominalism*. Oxford: Oxford University Press.
- . 2006. *Tracking reason: Proof, consequence and truth*. Oxford: Oxford University Press.
- . 2007. The inconsistency of natural languages: How we live with it. *Inquiry* 50(6): 590–605.
- . 2008. Alternative logics and the role of truth in the interpretation of languages. In *New essays on Tarski and philosophy*, ed. Douglas Patterson, 390–429. Oxford: Oxford University Press.
- . 2009. Why do informal proofs conform to formal norms? *Foundations of Science* 14: 9–26.
- . 2010a. The rule-following paradox and the impossibility of private rule -following. *The Baltic International Yearbook of Cognition, Logic and Communication. Volume 5: Meaning, understanding and knowledge*. doi:10.4148/biyclc.v5i0.278.
- . 2010b. *Talking about nothing: Numbers, hallucinations and fictions*. Oxford: Oxford University Press.
- . 2010c. Ontology and the word ‘exist’: Uneasy relations. *Philosophia Mathematica* (III) 18: 74–101.
- . 2012a. Responses to Gabriele Contessa, Erin Eaker, and Nikk Effingham. *Analysis* 72(2): 366–379.
- . 2012b. Referring to what is and what isn’t. In *Reference and referring*, ed. William P. Kabasenche, Michael O’Rourke, and Matthew H. Slater, 253–269. Cambridge, MA: MIT Press.
- . 2012c. Simple metaphysics and “ontological dependence”. In *Metaphysical grounding: Understanding the structure of reality*, ed. Fabrice Correia and Benjamin Schnieder, 234–253. Cambridge: Cambridge University Press.
- . 2013. *Semantic perception: How the illusion of a common language arises and persists*. Oxford: Oxford University Press.
- . forthcoming. Deflationist truth. In *Handbook on Truth*, ed. Michael Glanzberg. London: Blackwell.

- Baddeley, Alan, Michael W. Eysenck, and Michael C. Anderson. 2015. *Memory*. 2nd ed. London/New York: Psychology Press.
- Bayne, Tim, and Michelle Montague. (ed.). 2011. *Cognitive phenomenology*. Oxford: Oxford University Press.
- Blackburn, Simon. 1984. The individual strikes back. *Synthese* 58(3): 281–301.
- Boghossian, Paul. A. 1989. The rule-following considerations. *Mind* 98(392): 507–49.
- . 2003. The normativity of content. In *Philosophical Issues*, ed. Ernest Sosa and Enrique Villanueva, vol. 13: 31–45. Oxford: Blackwell Publishing.
- Boyd, Richard. 1991. On the current status of scientific realism. In *The philosophy of science*, ed. Richard Boyd, Philip Gasper, and J.D. Trout, 195–222. Cambridge, MA: The MIT Press.
- Butterworth, Brian. 1999. *The mathematical brain*. London: Papermac.
- Carey, Susan. 2009. Where our number concepts come from. *Journal of Philosophy* 106(4): 220–254.
- Carnap, R. 1956. Empiricism, semantics, and ontology. In *Meaning and necessity: A study in semantics and modal logic*, 205–221. Chicago: University of Chicago Press.
- Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Correia, Fabrice, and Benjamin Schnieder. 2012. *Metaphysical grounding: Understanding the structure of reality*. Cambridge: Cambridge University Press.
- Davidson, Donald. 1973. In defense of convention T. In *Inquiries into truth & interpretation* (1984), ed. Donald Davidson. Oxford: Oxford University Press.
- Dehaene, Stanislas. 1997. *The number sense*. Oxford: Oxford University Press.
- Descartes, René. 1993. *Meditations on first philosophy*. Trans. Donald A. Cress. Indianapolis/Cambridge: Hackett Publishing Company.
- Dorr, Cian, and John Hawthorne. 2013. Naturalness. In *Oxford studies in metaphysics*, ed. Karen Bennett and Dean W. Zimmerman, Vol. 8, 3–77. Oxford: Oxford University Press.
- Dupré, John. 1981. Natural kinds and biological taxa. *Philosophical Review* 90: 66–90.
- Fodor, Jerry A. 1974. Special sciences, or the disunity of sciences as a working hypothesis. In *Representations* (1981), 127–45. Cambridge, MA: MIT Press.
- . 1975. *The language of thought*. Cambridge, MA: Harvard University Press.
- . 1990. A theory of content, II: The theory. In *A theory of content and other essays*, 89–136. Cambridge, MA: The MIT Press.
- Forbes, Graeme. 1983–1984. Skepticism and semantic knowledge. *Proceedings of the Aristotelian Society, New Series*, vol. 8, 223–237.
- Fox, Douglas. 2011. The limits of intelligence. *Scientific American*, July, 36–43.
- Gelman, R., and C.R. Gallistel. 1986. *The child's understanding of number*. Cambridge: Harvard University Press (originally published in 1978).
- Gibbard, Alan. 2003. Thoughts and norms. In *Philosophical Issues*, ed. Ernest Sosa and Enrique Villanueva, Vol. 13, 83–98. Oxford: Blackwell Publishing.
- Ginsborg, Hannah. 2011. Primitive normativity and skepticism about rules. *The Journal of Philosophy* 108(5): 227–254.
- Glüer, Kathrin. 1999. Sense and prescriptivity. *Acta Analytica* 14: 111–128.
- Goldfarb, Warren. 1985. Kripke on Wittgenstein on Rules. *The Journal of Philosophy* 82(9): 472–488.
- . 1992. Wittgenstein on understanding. *Midwest studies in philosophy*, vol. 17, 109–122. Notre Dame: University of Notre Dame Press.
- Higginbotham, J. 1990. Contexts, model, and meanings: A note on the data of semantics. In *Mental representations: The interface between language and reality*, ed. R. Kempson, 29–48. Cambridge: Cambridge University Press.
- Hodes, Harold T. 1984. Logicism and the ontological commitments of arithmetic. *Journal of Philosophy* 81: 123–149.
- Hume, David. 1961. *A treatise of human nature*. New York: Doubleday & Company, Inc..
- Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.



- Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper & Row.
- . 1981. *Language & other abstract objects*. Totowa: Rowman and Little -field.
- Katz, Jerrold J., and Jerry A. Fodor. 1963. The structure of a semantic theory. *Language* 39: 170–210.
- Kim, Jaegwon. 1993. Multiple realization and the metaphysics of reduction. In *Supervenience and mind: Selected philosophical essays*, ed. Jaegwon Kim, 309–335. Cambridge, MA: Cambridge University Press.
- Kripke, Saul. 1982. *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.
- . 2013. *Reference and existence: The John Locke lectures*. Oxford: Oxford University Press.
- LaPorte, J. 1996. Chemical kind term reference and the discovery of essence. *Noûs* 30: 112–132.
- Lepore, Ernest. 1983. What model theoretic semantics cannot do. *Synthese* 54: 167–187.
- Lepore, Ernest, and K. Ludwig. 2000. The semantics and pragmatics of complex demonstratives. *Mind* 109(434): 199–240.
- Lewis, David. 1970. General semantics. In *Philosophical papers*, vol. 1 (1983), 189–229. Oxford: Oxford University Press.
- . 1983. New work for a theory of universals. In *Papers in Metaphysics and Epistemology* (1999), 8–55. Cambridge: Cambridge University Press.
- . 1984. Putnam's paradox. In *Papers in Metaphysics and Epistemology* (1999), 56–77. Cambridge: Cambridge University Press.
- Ludlow, P. 1999. *Semantics, tense, and time: An essay in the metaphysics of natural language*. Cambridge, MA: MIT Press.
- Mandler, G., and B.J. Shebo. 1982. Sublizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, 11–22.
- McDowell, John. 1984. Wittgenstein on following a rule. *Synthese* 58: 325–363.
- Moore, G.E. 1962. Four forms of scepticism. In *Philosophical papers*, 193–222. New York: Collier Books.
- Pietroski, Paul M. 2003. The character of natural language semantics. In *Epistemology of language*, ed. Alex Barber, 217–256. Oxford: Oxford University Press.
- Pitt, David. 2004. The phenomenology of cognition, or, what is it like to think that P? *Philosophy and Phenomenological Research* 69: 1–36.
- Priest, Graham. 2011. Review of “Talking about Nothing”. *Philosophia Mathematica (III)* 19: 359–363.
- Putnam, Hilary. 1967. The nature of mental states. In *Mind, language and reality: Philosophical papers*, vol. 2 (1975), 429–40. Cambridge: Cambridge University Press.
- . 1978. Meaning and knowledge. In *Meaning and the moral sciences*, ed. Hilary Putnam, 7–80. London: Routledge & Kegan Paul.
- . 1981. *Reason, truth and history*. Cambridge: Cambridge University Press.
- Quine, W.V. 1981. Responses. In *Theories and things*, ed. W.V. Quine, 173–186. Cambridge, Massachusetts: Harvard University Press.
- Rosch, E., and C.B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605.
- Rosch, E., C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382–439.
- Sider, Theodore. 2009. Ontological realism. In *Metametaphysics: New essays on the foundations of ontology*, ed. David J. Chalmers, David Manley, and Ryan Wasserman, 384–423. Oxford: Oxford University Press.
- . 2011. *Writing the book of the world*. Oxford: Oxford University Press.
- Strawson, Galen. 1994. *Mental reality*. Cambridge, MA: MIT Press.
- Sundell, Timothy. 2012. Disagreement, error, and an alternative to reference magnetism. *Australasian Journal of Philosophy* 90(4): 743–759.

- Tait, William. 1986. Wittgenstein and the “Skeptical Paradoxes,” reprinted in (William Tait) *The provenance of pure reason*, 198–211 (2005). Oxford: Oxford University Press.
- Walker, Ralph C.S. 1989. *The coherence theory of truth: Realism, anti-realism, and idealism*. London: Routledge.
- Wikforss, Åsa Maria. 2001. Semantic normativity. *Philosophical Studies* 102(2): 203–226.
- Williams, J. Robert G. 2007. Eligibility and inscrutability. *Philosophical Review* 116(3): 361–399.
- . 2015. Lewis on reference and eligibility. In *A companion to David Lewis*, ed. Barry Loewer and Jonathan Schaffer, 367–381. Malden: Wiley Blackwell.
- Williamson, Timothy. 1994. *Vagueness*. London: Routledge.
- Wilson, Mark. 1982. Predicate meets property. *Philosophical Review* 41: 549–589.
- Wittgenstein, Ludwig. 1956. *Remarks on the foundations of mathematics*. Oxford: Basil Blackwell.
- . 1958. *Philosophical Investigations*. Trans. G.E.M. Anscombe, 3rd ed. New York: The Macmillan Company.
- Wright, Crispin. 1980. *Wittgenstein on the foundations of mathematics*. London: Duckworth.
- . 1984. Kripke’s account of the argument against private language. *The Journal of Philosophy* 81(12): 759–778.
- . 1989. Critical notice of Colin McGinn’s Wittgenstein on meaning. *Mind* 98: 289–305. Reprinted in (Alexander Miller and Crispin Wright) *Rule-following and meaning*, 108–128. Montreal/Kingston: McGill-Queen’s University Press. References are to the reprint.