ACTA PHILOSOPHICA GOTHOBURGENSIA
ISSN 0283-2380

Editors:
Helge Malmgren, Christian Munthe, Ingmar Persson and Dag Westerståhl

Published by the Department of Philosophy of the University of Göteborg

Subscription to the series and orders for single volumes should be addressed to:
ACTA UNIVERSITATIS GOTHOBURGENSIS
Box 222, SE-405 30 Göteborg, Sweden

VOLUMES PUBLISHED

12. LARS SANDMAN: A Good Death: On the Value of Death and Dying. 2001. 346 pp.

13. KENT GUSTAVSSON: Emergent Consciousness. Themes in C.D. Broad's Philosophy of Mind. 2002. 204 pp.

14. FRANK LORENTZON: Fri vilja? 2002. 175 pp.

15. JAN LIF: Can a Consequentialist Be a Real Friend? (Who Cares?). 2003. 167 pp.

16. FREDRIK SUNDQUIST: Perceptual Dynamics. Theoretical Foundations and Philosophical Implications of Gestalt Psychology. 2003. 261 pp.

17. JONAS GREN: Applying Utilitarianism. The Problem of Practical Action-guidance. 2004. 160 pp.

18. NIKLAS JUTH: Genetic Information – Values and Rights. The Morality of Presymptomatic Genetic Testing. 2005. 459 pp.

19. SUSANNA RADOVIC: Introspecting Representations. 2005. 200 pp.

20. PETRA ANDERSSON: Humanity and Nature. Towards a Consistent Holistic Environmental Ethics. 2007. 190 pp.

21. JAN ALMÄNG: Intentionality and Intersubjectivity. 2007. 210 pp.

22. ALEXANDER ALMÉR: Naturalising Intentionality. Inquiries into Realism & Relativism. 2007. 284 pp.

23. KRISTOFFER AHLSTRÖM: Constructive Analysis. A Study in Epistemological Methodology. 2007. 308 pp.

ISBN 978-91-7346-603-5

KRISTOFFER AHLSTRÖM CONSTRUCTIVE ANALYSIS

Acta Philosophica Gothoburgensia 23

CONSTRUCTIVE ANALYSIS

A Study in Epistemological Methodology

KRISTOFFER AHLSTRÖM

GÖTEBORG UNIVERSITY
Acta Universitatis Gothoburgensis

Constructive Analysis

A Study in Epistemological Methodology

Acta Philosophica Gothoburgensia 23

CONSTRUCTIVE ANALYSIS

A Study in Epistemological Methodology

KRISTOFFER AHLSTRÖM

GÖTEBORG UNIVERSITY
Acta Universitatis Gothoburgensis

Distribution:
ACTA UNIVERSITATIS GOTHOBURGENSIS
Box 222
SE-405 30 Göteborg
Sweden

Typeface: Garamond 10/14 pt.
Cover design: Peter Johnsen

ISBN 978-91-7346-603-5
ISSN 0283-2380
Printed by Geson, Göteborg 2007

If an epistemological theory tells us that a particular policy of belief formation is justified or a particular type of inference is rational, and that these claims are analytic, that they unfold our concepts of justification and rationality, an appropriate challenge is always, "But why should we care about these concepts of justification and rationality?" The root issue will always be whether the methods recommended by the theory are well adapted for the attainment of our epistemic ends, and that cannot be settled by simply appealing to our current concepts.

— PHILIP KITCHER

Sometimes the purposes of explanation and understanding are best served by not talking the way our grandparents talked.

— FRED DRETSKE

Contents

Acknowledgements

Before proceeding to the main text, I would like to extend a few thanks. First of all, I would like to thank my (current and previous) advisers Helge Malmgren, Åsa Wikforss, and Roland Poirier Martinsson, as well as Erik J. Olsson for constant, unyielding support and many hours put into making sense of the questions and issues addressed in the present study.

I would also like to thank the faculty of New York University's Department of Philosophy—Stephen Schiffer and Don Garrett, in particular—for providing me with the opportunity to spend two semesters and one summer in an exceptionally inspiring and knowledgeable environment, first as a Fulbright Fellow and then as a Visiting Scholar, all the while working out some of the basic premises of the present study. On that note, I would also like to extend my deep gratitude to the Fulbright Program and, in particular, to the wonderful and always so helpful staff at the Swedish Fulbright Commission in Stockholm.

I am also extremely grateful to Jubileumsfonden at Göteborg University for providing me with a generous grant that enabled me to spend six months at University of Massachusetts' Department of Philosophy in Amherst, where the majority of the present manuscript was finished and parts presented. I would also like to express my deep

appreciation for the entire faculty and graduate student body there for making my stay such a pleasant and stimulating experience. In particular, I would like to thank Fred Feldman for being so helpful in the process of setting everything up, and Hilary Kornblith for untiring encouragement, apt guidance, and several hours of invaluable discussion on many of the issues on which the present study turns.

Needless to say, many more have, at one point or another, contributed to the improvement of the material at hand (while being in no way responsible for any faults that it may still have). Apart from the ones already mentioned, I would like to take this opportunity to also thank Louise Antony, Jan Almäng, Arvid Båve, Jeremy Cushing, Sinan Dogramaci, Ragnar Francén, Kent Gustavsson, Angus Hawkins, Peter Johnsen, Klemens Kappel, Felix Larsson, Joe Levine, Anna-Sara Malmgren, Gareth Matthews, Kirk Michaelian, Sven Nyholm, Alex Sarch, Nico Silins, Matthew Smith, Anders Tolland, and Dag Westerståhl.

Last, but in no way least, I would like to thank Radha—my heart, my happiness, *mi vida*. Words ring hollow when I try to say how much you mean to me. But if I'm doing anything right, words won't be necessary; you will know that you're my everything.

Northampton, MA, December 2007

Typography and Abbreviations

To avoid confusion, the following typographical principles will be adhered to in the present study, unless otherwise is stated:

> *Concepts*: Words or expressions in capital letters (e.g., HORSE or SPATIAL LOCATION) designate concepts. Capital letters surrounded by single quotation marks (e.g., '*F*') serve as concept variables.
>
> *Linguistic terms*: Words or expressions surrounded by double quotation marks (e.g., "horse" or "spatial location") designate linguistic terms or, better said, the mention rather than the use of the expressions in question. Capital letters surrounded by double quotation marks (e.g., "*F*") serve as linguistic term variables.
>
> *Referents*: Words or expressions that are *not* surrounded by any quotation marks (e.g., horse or spatial location) designate the objects or phenomena picked out by the corresponding term or concept. Capital letters surrounded by no quotation marks (e.g., *F*) serve as referent variables.

For the sake of brevity, I will also use a series of abbreviations, listed below for the convenience of the reader:

AC	The Accessibility Constraint
CA	Constructive Analysis
CIT	The Cognitive-Introspective Thesis
CPA	Conceptual Purpose Analysis
CPLM	Clinical Prediction via Linear Models
DCA	Definitional Conceptual Analysis
DF	Diagnosis via Frequencies
FA	Factual Analysis
GC	The Guidance Conception of Epistemology
IBAI	Introspection-Based Access Internalism
MA	Meaning Analysis
NRE	Analysis via Narrow Reflective Equilibrium
PCA	Prototypical Conceptual Analysis
RC	A Reconstructed Concept of Justification
RE-F	Retention through Elaboration—Frequentist
RE-G	Retention through Elaboration—General
REL	Retention through Elaboration and Lag
S	Epistemic subject (variable)
SPR	Statistical Prediction Rules
SST	Selection via Statistical Training
WRE	Analysis via Wide Reflective Equilibrium

Introduction

Epistemology has some serious work to do. Surely, no epistemologist would deny that. Still, it is one of the main tenets of the present study that much of contemporary epistemology has not been conducted in the way that it should. More specifically, it has not been conducted in a way conducive to what should be one of its main goals, namely to guide epistemic inquiry in the attainment of our most central epistemic goals. Furthermore, it will be claimed that the very reason that epistemology has failed to do so pertains to what I will argue is an entrenched—indeed, in a sense, a literally *ancient*—but implausible methodology. In fact, I will not only (*a*) argue that we ought to revise this methodology and (*b*) put forward an alternative, but also (*c*) demonstrate the usefulness of this alternative methodology within the analysis of epistemic justification. There, it will first be argued that some of the most influential theories fail and that our concept of justification—considering the goals of epistemic inquiry—is best *reconstructed* in terms of truth-conductivity, and then, through a discussion of psychological research relevant to reasoning strategies, shown how such a reconstructed concept may be used to improve on actual truth-seeking inquiry.

More specifically, the structure of the present study is as follows. Part I lays out and criticizes the entrenched methodology as well as presents and defends an alternative. The latter methodology is then applied in Part II in relation to the epistemological discussions about epistemic justification. Part I is, in turn, sub-divided into four chapters. In chapter 1, I argue that there is a striking continuity between Plato's Socratic method and the methodology of contemporary philosophy, in that both Platonic and contemporary analysis is best construed as a pursuit of *definitions* by way of *intuitions*. Although this methodology might very well have made sense to Plato, it is not so clear that it is well-motivated when divorced from an ontology of Forms and a faculty of rational insight. In particular, I argue that there is little support for the claim that concepts—i.e., what much of contemporary philosophy takes to be its proper objects of study—are best characterized in terms of what has become the format of choice: simple, clear-cut, necessary, and sufficient conditions.

Chapter 2 considers a rectification in terms of so-called prototypes, which some contemporary psychologists take to be a more suitable model for representing concepts. However, attending to psychological evidence thus only serves to raise the question whether a traditional armchair method at all presents a sensible methodology in the characterization of concepts, in light of the substantially more rigorous methods of empirical psychology. I will argue for a negative answer. But I will also note that the question becomes relevant only under the assumption that we should be at all interested in exhaustive accounts of our concepts. *Contra* this assumption, I suggest that our (present) concepts, at best, make up an interesting domain of epistemological study by providing the preliminary material in the search for an epistemic vocabulary that serves us *better*—a task that does not require anything like exhaustive accounts of our concepts.

This is, to some extent, appreciated by philosophers characterizing their inquiry as an attempt to reach a reflective equilibrium between their concepts, general norms, and (in some cases) best empirical theories about the world. However, the exact details of how to reach such an equilibrium are, unfortunately, not all that clear, beyond

the valid but vague point that concepts and norms should not be insulated from external scrutiny.

This idea is further elaborated on in chapter 3, in relation to a recent suggestion by Hilary Kornblith (e.g., 2006, 2002) to the effect that epistemology, if not philosophy at large, is a substantially empirical inquiry. I show that his point is, in fact, not contingent upon the further and admittedly controversial claim that all objects of investigation are natural kinds, but is perfectly compatible with all such objects being artifactual kinds. However, while granting that a lot of philosophical investigation, thereby, should proceed in such a way that concepts (and the analysis thereof) take the back seat as soon as they have served to fix the subject matter, I argue that this cannot be the end of the methodological story. Picking up the thread from chapter 2, I make a case for the claim that, to the extent that we are concerned with *normative* phenomena and concepts—which is often the case in epistemology—it makes sense to not only investigate referents and *refine* our concepts to the extent that they do not provide an accurate story of the phenomena they refer to, but also ask whether it is possible to *reconstruct* a concept that serves us better in light of what we might find out not only about its referent but also about its purpose.

In chapter 4, I recapitulate the lessons from the previous chapters by identifying two components to a revised methodological framework—one descriptive and one ameliorative. The first component consists in the two-fold descriptive task of (*a*) *identifying* the subject matter through an elucidation of the relevant epistemic concept(s) and (*b*) *aggregating* a characterization of the phenomenon referred to by this (or these) concept(s). While the latter is best carried out by way of fairly straightforward empirical investigation, the former will turn out to call for something reminiscent of, albeit far less demanding than, conceptual analysis, as traditionally construed. More specifically, I argue for a notion of *meaning analysis*, building upon Hilary Putnam's theory about meaning and *stereotypes*.[1] While neither (necessarily) providing an accurate picture of the phenomenon to which they refer,

[1] See Putnam (1975a, b).

presupposing a substantial distinction between analytic and synthetic truths, nor determining the corresponding terms' reference, such stereotypes play a non-trivial role in understanding and communication and constitute a powerful, cognitive pathway to extra-mental phenomena. Hence, stereotypes demarcate a useful notion of meaning.

The second methodological component consists in the ameliorative task of (*a*) *evaluating* the extent to which our current epistemic concepts serve their purposes and (*b*) *improving* our concepts to the extent that they do not. Again, traditional conceptual analysis does not seem to get us what we need here. I will argue that a more plausible candidate for the job is (conceptual) *purpose analysis*—a kind of analysis that builds upon some recent suggestions from Jonathan Weinberg and Edward Craig.[2] As for improving our epistemic concepts, I suggest that improvement should be conducted by way of further empirical investigations—preceded and guided by a proper purpose analysis—into candidate characteristics that, on incorporation, would enable us to attain our epistemic goals to a greater degree, where our epistemic goals are to be understood in relation to the attainment and maintenance of true belief in significant matters.

I will refer to this revised framework of analysis as *constructive* analysis—a name chosen because of its dual connotation to the refinement and *reconstruction* of epistemic concepts as well as to the idea of, thereby, *serving a useful purpose* for actual epistemic inquiry in naturalistic settings. This brings us to Part II of the study, where constructive analysis is applied to epistemic justification—a phenomenon that has been subject to a lot of philosophical debate at least since Edmund Gettier's famous critique of the classic tripartite analysis of knowledge.[3] In chapters 5 and 6, I consider two suggestions as to how we may understand this notion: epistemological deontologism and introspection-based access internalism. I argue that adherents of the former face a dilemma: either they opt for a substantial notion of

[2] See Weinberg (2006) and Craig (1990)

[3] See Gettier (1963).

intellectual duty, in which case they, under the principle that 'ought' implies 'can,' have to commit themselves to the highly implausible thesis that we may chose voluntarily what to believe; or they resort to merely talking about duties in a metaphorical sense and, thereby, leave behind the very normative discourse that originally motivated their position. That is, unless they opt for a reconstruction along the lines of introspection-based access internalism, i.e., the idea that justification pertains not to intellectual duties, but to the sound evaluation of our epistemic reasons by way of introspection, understood as the notice that the mind takes of its own states and operations. However, in light of recent evidence within cognitive psychology, the claim that we in general have anything like a reliable (or powerful) introspective access to the reasons actually underlying our beliefs seems highly dubious.

This calls for a more plausible reconstruction. In chapter 7, I turn to my own attempt, taking into account not only what we have found out about the qualities that have (for better or worse) been taken to pertain to justification, but also the specific purpose that the corresponding concept can reasonably be expected to fill, in light of the norms in which it figures and the goals that endow it with normative force. Starting out with the quite pervasive idea that being justified pertains to having good reasons for taking a (set of) proposition(s) to be true, I venture into the details of the specific purpose JUSTIFICATION plays in relation to the goals identified in chapter 3. I conclude that the purpose of JUSTIFICATION, at the most general level, is to flag appropriate sources of information within the various (social) practices involved in the exchange of information. In light of the failure of deontologism and introspection-based access internalism, however, I argue that the most plausible way to spell out this appropriateness is in terms of *effective heuristics*, designating justification conferring processes on a continuum stretching from basic belief-forming processes to conscious reasoning strategies, effective in so far as they strike a good balance between generating a lot of true belief (power) and generating a majority of true belief (reliability).

An important—and, in my opinion, extremely welcome—consequence of such a concept of justification, is that it identifies justification as a perfectly natural phenomenon, pertaining to the actual and possible track records of belief-forming mechanisms and strategies. In fact, as I will argue in chapter 8, this concept also enables us to improve our cognitive outlooks, when combined with what cognitive science and psychology may teach us about the ways in which we tend to reason. More specifically, focusing on the three central epistemic endeavors *prediction*, *diagnosis*, and *retention*, and armed with a concept that speaks to our epistemic goals, it is demonstrated how the reconstructed concept of justification may be implemented within and improve on actual truth-seeking inquiry in naturalistic settings, through the advancement of sound reasoning strategies. Thereby, it is also shown how epistemology, conducted in terms of constructive analysis, may live up to the noble and time-honored epistemological challenge of not only describing but also guiding epistemic inquiry—a challenge that ought to lie at the heart of any naturalistic epistemology.

Part I.

Philosophical Analysis and Epistemological Methodology

Chapter 1.
From Forms to Concepts

Shortly before his trial and execution in 399 BC, Socrates meets with Theodorus—a skilled geometrician from Cyrene. As Plato tells the story, Theodorus introduces Socrates to one of his students, Theaetetus, with whom Socrates soon engages in a discussion. The topic is the nature of knowledge, which turns out to be quite elusive. Theaetetus is able to pin down a couple of paradigm examples, but laments that it is so much more difficult to find a satisfactory definition of knowledge than it is to define the properties that he encounters in class with Theodorus. Socrates encourages him to keep trying:

> *Socrates*: Come, you made a good beginning just now; let your own answer about roots be your model, and as you comprehended them all in one class, try and bring the many sorts of knowledge under one definition.
>
> *Theaetetus*: I can assure you, Socrates, that I have tried very often, when the report of questions asked by you was brought to me; but I can neither persuade myself that I have a satisfactory answer to give, nor hear of any one who answers as you would have him; and I cannot shake off a feeling of anxiety.

> *Socrates*: These are the pangs of labour, my dear Theaetetus; you have something within you which you are bringing to the birth.[1]

In a dialogue that came to bear his name, Theaetetus takes Socrates' encouragement to heart and embarks upon an excellent example of what has come to be known as *the Socratic method*. On this method, questions of the form "What is (an) *F*?" are approached through suggested definitions, disqualified in so far as they fail to include all intuitive instances (things that intuitively are *F*) or exclude unintuitive instances (things that intuitively are *not F*). And, only a definition that survives scrutiny for such counterexamples constitutes "a noble and true birth."[2]

1.1. Whence the Socratic Method?

When attempting to understand this method, it is important to note what Socrates did *not* ask Theaetetus to do: He did not say, "Theaetetus, waste no more time talking to me—go examine actual instances of knowledge and then come back and tell me what knowledge is." No, Socrates tells Theaetetus to look *inwards*, to attend to that which he is "bringing to the birth," in a persistent examination (and quite frequent refutation) of suggested *definitions* through the probing of *intuitions*. Why did Socrates do this? On the face of it, it seems that this, at best, would reveal what Theaetetus *believes* about knowledge—not what knowledge really *is*. Still, Socrates is, clearly, interested in the latter. So, why does he ask Theaetetus to look inwards? The reason may be brought out as follows.[3]

---

[1] The *Theaetetus* in Plato (1953, p. 148de).

[2] The *Theaetetus* in Plato (1953, p. 150c).

[3] The following interpretation of such a rich and subtle body of work as Plato's makes no claim of being the only possible one. My aim here is merely to deliver an interpretation that makes sense of Plato's notion of analysis,

First, consider Plato's ontology: The world is split into two ontologically distinct domains. Within the first domain, we find the fleeting phenomena encountered in ordinary sense experience, at times making up quite a disparate collection of things and events. In the second domain, we find a set of perfect and immutable *Forms*, serving as *universals*. A universal is that which is predicable of many, such as *Redness* and *Roundness*, and, hence, can be instantiated by numerically distinct entities. Plato famously takes these Forms or universals to be ontologically separate from their instantiations. That is, even if there were no red things, there would still be Redness (i.e., the Form of Redness)—an idea that his student Aristotle later came to criticize. More specifically, Plato believes that

(1.1.1) instantiating a general property is a matter of *taking part in* the immutable Form corresponding to that property.

Furthermore, the fact that instantiations take part in immutable Forms serves to *unite* the fleeting and imperfect phenomena of the first domain and, in effect, explains why particular yet distinct things (this red saucer, that red ball) may, nevertheless, fall in identical categories (red, round), since

(1.1.2) every Form has an *essence*, determining the fundamental nature of its instantiations.

In fact, Plato seems to think not only that every Form has an essence, uniting its instantiations, but also that there is a certain sentential structure that is particularly suited for capturing such essences. More specifically,

---

especially as it relates to modern philosophical practice. Needless to say, attempts to make sense of other aspects of Plato's philosophy might very well yield different yet—given different ambitions—equally reasonable interpretations.

(1.1.3) every essence is specifiable through *a necessary and sufficient condition* for taking part in the corresponding Form.

Given that the format best suited for providing necessary and sufficient conditions is that of a definition, we may conclude that,

(A_F) for every Form *F*—or, at least, most philosophically interesting Forms—there is a *definition*, specifying the necessary and sufficient condition for taking part in *F*.

It is in this context that we find Socrates insisting on bringing all the various instances of knowledge under one definition—the linguistic counterpart of an essence. As such, it is not just any definition but a *real* definition, concerned with essences and real natures (in contemporary philosophy primarily pertaining to natural kinds, as we shall see below) rather than the meanings of words, or what may be called a *nominal* definition.

Still, this does not explain why Socrates would insist on generating such definitions via the probing of *intuitions*. To explain this, we need to consider Plato's semantics and epistemology. Over and above postulating Forms, Plato also assumes that there is a correlation between the Forms and the sets of things we tend to give the same name, such that

(1.1.4) every general term corresponds to a Form.

This is Plato's famous statement of the One-Over-Many Principle, as expressed in the *Republic*.[4] Furthermore, he seems to assume that there

[4] See the *Republic* in Plato (1953, p. X 596a). In his later dialogues, Plato is more skeptical about the extent to which our conceptual apparatus cuts the world at its joints. See, e.g., the *Sophist* in Plato (1953, p. 218bc) and the *Statesman* in Plato (1953, p. 262d-263a).

is a connection between general terms and Forms that enables us to go from the competent use of the former to at least the beginning of a grasp of the latter. To see this, consider the so-called Paradox of Inquiry, later to resurface as the Paradox of Analysis in C. H. Langford's discussion of G. E. Moore's notion of analysis,[5] but originally posed by Meno (in the Platonic dialogue bearing his name) in relation to the following two questions:

> If you already know what *F* is, how can there be a genuine search as to the nature of *F*?
>
> If you do *not* know what *F* is, how will you know (*a*) how to aim your search or (*b*) that you have stumbled upon a satisfactory account of *F*?

One way to understand these questions is as posing a challenge to identify a middle-road between a *complete* insight into *F* (making analysis redundant) and a complete *lack* of insight into *F* (making analysis impossible). In the *Meno*, Socrates' direct response is the Doctrine of Recollection. According to this doctrine—constituting the culmination of Socrates' famous interrogation of the slave boy about how to double the area of a square—learning in general and analysis in particular is a process involving a recollection of something that we used to know when our souls were still disembodied. Again, it is helpful to construe the invocation of this admittedly somewhat obscure doctrine in terms of an ambition to pinpoint an *incomplete* grasp, such that it is incomplete enough to not make analysis redundant, yet complete enough to not make it impossible. In this context, it is interesting to note the question Socrates asks Meno about the slave boy before initiating the interrogation: "He is a Greek and speaks our language?"[6] Why would Socrates ask that? Well, supposedly, he wants to be able

---

[5] See Langford (1942).

[6] The *Meno* in Plato (1970, p. 82b). Thanks to Gareth Matthews for directing my attention to this passage.

to communicate with the boy and a prerequisite for successful communication is speaking the same language. Hence, when Socrates asks the boy if he knows what a square is, the boy has some idea of what he is talking about. And the same goes for "line," "equal," "larger," "smaller," and so on, for all the terms Socrates uses in his discussion with the boy. More specifically, it is reasonable to assume that, in the general case,

> (1.1.5) competently *employing* a general term "*F*" at the very least involves having an *incomplete grasp* of the corresponding Form *F*, in the sense of being able to (*a*) realize of some instances that they are instances of *F* and (*b*) realize of some other phenomena that they are not instances of *F*.

This, furthermore, seems to be the strategy of Plato when he approaches the problem in the *Phaedo*, where he has Socrates indicate that it is possible to recognize (at least some) positive as well as negative instances of a Form, without already being in the possession of a definition, specifying the necessary and sufficient condition for instantiation.[7] More specifically, assuming (1.1.5), we may approach the paradox as follows: The circumstance in which an analyzer finds herself is such that she does *not* already possess the necessary and sufficient condition specifying the essence of the analysandum Form. Yet, she is, *qua* competent user of the term "*F*," able to identify some instances as instances of *F* and some other phenomena as *not* being instances of *F*. Thereby, she has both (*a*) the material to aim her search for a complete account of *F*, and (*b*) an idea of what such an account would look like (where this idea goes beyond mere formal constraints such as non-circularity, etc.).

Having delimited the circumstance of the analyzer in terms of a state of incomplete grasp or insight, we may understand the goal of

---

[7] See the *Phaedo* in Plato (1953, p. 100de).

analysis as that of *fully* grasping the essence of the analyzed Form. Given (1.1.3), we may spell this out in such a way that

> (1.1.6) *fully* grasping the essence of a Form consists in grasping the necessary and sufficient condition for taking part in that Form.

Now, Plato makes the epistemological assumption that the proper way to gain a more complete insight into Forms—that is, the proper way to go from the incomplete grasp involved in competently employing the term to a full grasp of the corresponding Form—is through the probing of *intuitions*, i.e., our dispositions to categorize entities as being an instance of a particular Form—be it the Form of temperance, courage, piety, virtue, knowledge, etc. Why would Plato make such an assumption?

The most obvious Platonic rationale invokes, again, the Doctrine of Recollection, now understood as the idea that we (*a*) are in possession of true beliefs about the essences of Forms, (*b*) fail to know that we are due to an inevitable forgetfulness that takes place prior to becoming embodied, yet (*c*) may still successfully convey parts of these true beliefs through our ability to competently use general terms. More specifically, Socrates suggests—right after having put forward the Doctrine of Recollection—that the way to *access* these forgotten insights is not through teaching but *questioning*, since only the latter enables one "to discover—that is, to recollect—what one doesn't happen to know or (more correctly) remember, at the moment."[8]

This is why the spontaneous judgments of competent speakers provide the basic material for the Socratic method, and probably also why Plato chose to present his views in the form of dialogues; in a process involving careful scrutiny, questioning and answering, these judgments might just be carrying important information about the Forms. In other words,

---

[8] The *Meno* in Plato (1970, p. 86b).

(B_F) the probing of intuitions may serve to *elucidate* the conditions defining the essences of Forms.[9]

Consequently, we find Socrates referring to himself as a midwife, encouraging Theaetetus to look inwards—not outwards—and attend to that which he is bringing to birth, in a constant probing of intuitions through repeated questioning and refutation in light of hypothetical cases. Because under the assumption that there is an immutable Form for knowledge, all the imperfect instances of knowledge within the domain of everyday life are perfectly uninteresting, as far as true understanding goes. What is *not* uninteresting, however, is that (*a*) all instances of knowledge have an *essence* by virtue of taking part in a single immutable Form of knowledge—an essence specifiable through a *definition* in terms of a necessary and sufficient condition for taking part in this Form—and (*b*) it is possible to *elucidate* this essence and, hence, make available the relevant definition, by the appropriate probing of our intuitions. Hence, the Socratic method.

Now, if (A_F) and (B_F) were nothing but revered pieces of an outdated philosophical theory, Plato's discussion with Theaetetus would (justifiably) be reduced to historical anecdote. However, I will in the following two sections first point to a striking methodological similarity between the Platonic dialogues and modern analytical philosophy, and then argue that, even in spite of the fact that few (if any) philosophers today explicitly espouse Plato's ontology, semantics, and epistemology,[10] the best explanation of this similarity is that contemporary analogues of (A_F) and (B_F) implicitly underlie much of contemporary analytical philosophy.

---

[9] Cf. Ramsey (1998, p. 165).

[10] There is even a scholarly debate as to whether Plato still defends an ontology of Forms in the *Theaetetus*. See, e.g., McDowell (1973).

1.2. Platonic Remnants in Contemporary Epistemology

Turn now to contemporary epistemology, where the following may serve to illustrate the continuity in methodology: According to a traditionally influential epistemological analysis, knowledge is justified true belief.[11] In other words, if and only if a subject believes a true proposition, and has good reasons for doing so, she knows that proposition. And this is not just any claim about knowledge—it is a *definition* of knowledge. Consequently, and just like in the Platonic dialogues, epistemologists have come to evaluate it by determining whether it, on reflection, is susceptible to any intuitive counterexamples, i.e., whether it, in the words of Frank Jackson, survives "the method of possible cases."[12] So, consider the following scenario from a seminal paper by Edmund Gettier:

> Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:
>
> > (d) Jones is the man who will get the job, and Jones has ten coins in his pocket.
>
> Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails:
>
> > (e) The man who will get the job has ten coins in his pocket.

---

[11] The exact source of this analysis is not altogether clear. Indeed, it has even been suggested by Alvin Plantinga (1990, p. 45) that the "tradition" underlying the analysis might not be more than an artifact of the extensive critique it has had to endure. However, see Chisholm (1957) and Ayer (1956) for two oft-cited cited proponents.

[12] See Jackson (1998).

> Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.
>
> But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (*i*) (e) is true, (*ii*) Smith believes that (e) is true, and (*iii*) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.[13]

When confronted with examples of this kind, a lot of philosophers have found themselves inclined to agree with Gettier; intuitively, there are instances of justified, true belief that do not count as knowledge. What does this tell us about the proposed definition? As noted by James Cornman, Keith Lehrer, and George Pappas in *Philosophical Problems and Arguments: An Introduction*:

> […] we shall tentatively consider a definition satisfactory if, after careful reflection, we can think of no possible examples in which either the defined word truly applies to something but the defining words do not, or the defining words truly apply to something but the defined word does not. When we can think of such an example, then we have found a counterexample to the alleged definition showing that we do not have an accurate reportive definition. If we can find no counterexample to a

[13] Gettier (1963, pp. 121-122, emphasis in original).

> definition, then we may regard it as innocent until a counter-example is found to prove otherwise.[14]

Hence, according to entrenched philosophical dialectics, we conclude that knowledge is not justified true belief.[15] And, what we have just seen is nothing less than a paradigm example of epistemological analysis. Epistemological analysis, as it has traditionally been construed and is still carried out to a large extent, is the analysis of epistemic concepts via intuitions. However, the intuitions relied on are not just any intuitions. On a very liberal reading, an intuition is any fairly direct belief that is associated with a strong feeling of being true. This is the sense in which we may "intuit" that there is an external world and that nothing can be blue and green all over. What we will be concerned with here, however, is something different, namely what is typically referred to as *categorization intuitions*. Such intuitions take the following form:

> Phenomenon *x* is (not) an instance of (the concept) '*F*.'

Such categorization intuitions are taken to provide interesting, philosophical data by virtue of being caused by and, hence, sensitive to underlying *categorization dispositions*, in turn caused by the concepts that the subject in question possesses. In short, categorization intuitions reveal concepts, i.e., the very target of much contemporary epistemology. We will return to the details of the relation between concepts and intuitions below. For now, however, we only need to note that we have delimited the sense in which epistemological analysis is *conceptual analysis*.[16] And the same goes for analytical philosophy at large, which, in the methodological vein of Plato, is best described as the enterprise

---

[14] Cornman, Lehrer, and Pappas (1982, p. 18).

[15] For a penetrating and exhaustive survey of the first two decades of discussion following Gettier's paper, see Shope (1983).

[16] See Kornblith (2002). See also Stich (1998) on the prevalence of *analytical epistemology*.

of constructing theories—be it about knowledge, consciousness, the morally obligatory, etc.—the plausibility of which is taken to be largely, if not completely, determined by their susceptibility to counterexamples derived from categorization intuitions, supposedly uncovering the very concepts that the theories are supposed to capture.

It is important, however, to acknowledge the leeway between "largely" and "completely" in the previous sentence, considering that different philosophers committed to a conceptual analytic methodology might, nevertheless, bestow categorization intuition with different evidential value. Hence, Saul Kripke, one of the most influential philosophers of the 20th century:

> […] some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have in favor of anything, ultimately speaking.[17]

As Kripke evidentially is fully aware of, some philosophers take a more moderate standpoint. For example, according to David Lewis—one of the 20th century's most prominent metaphysicians—categorization intuitions certainly play an important role in philosophical theorizing, in supplying us with a set of "pre-philosophical opinions" that ought to be respected in so far as we are "firmly attached" to them. However, according to Lewis there is also a certain amount of "give-and-take" in the construction of philosophical theories as a result of the possibility of conflicts between respecting *all* our intuitions and providing a fully *systematic* account.[18] We will return to this idea in the next chapter when discussing so-called reflective equilibrium approaches to analysis.

Still, even on a moderate account, categorization intuitions play a central and important methodological role, and have done so

---

[17] Kripke (1980, p. 42).

[18] See Lewis (1973, p. 88).

for quite a while. As noted by Goldman, this is certainly not to say that philosophers have always described their methodology in the language of intuitions.[19] Take, for example, Locke's discussion of personal identity and the famous prince-cobbler case:

> For should the soul of a prince, carrying with it the consciousness of prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, *every one sees* he would be the same person with the prince […].[20]

Clearly, Locke's talk about what "every one sees" is easily translatable into what everyone intuits. In more recent years, philosophers have come to talk more explicitly in terms of intuitions. As the above made clear, the Gettier discussion constitutes an exceptionally clear example of contemporary, intuition-driven philosophy, as does the literature on causal theories of meaning after the publication of Putnam and Kripke and the discussion of personal identity as discussed by Derek Parfit and Judith Jarvis Thomson.[21] In fact, this practice of supporting and refuting philosophical analyses with reference to categorization intuitions and the concepts that they, supposedly, reveal, is so widespread that it has been referred to it as part of "the standard justificatory procedure" in philosophy.[22]

Hence, Michael DePaul and William Ramsey:

> Refutations by intuitive counterexamples figure as prominently in today's philosophical journals as they did in Plato's dialogues. In recent times, efforts to provide philosophical analyses of knowledge, the nature of meaning and reference, the human mind, and moral right and wrong—to name only a few examples—have been both defended and attacked by appeal

---

[19] See Goldman (2007).

[20] Locke (1996, book II, chapter xxvii, 15; my emphasis).

[21] See Putnam (1975b), Kripke (1980), Parfit (1984), and Thomson (1971).

[22] See Bealer (1998).

> to what is considered to be intuitively obvious. Even philosophers who do not advertise themselves as engaged in the search for necessary and sufficient conditions nevertheless lean heavily upon our judgments and counterexamples to support or criticize positions. While there have always been a few philosophers who have been skeptical of the search for precise analyses, this type of philosophy is still very widely practiced. For many, appealing to our intuitions is the only available option for uncovering the true nature of the many things that occupy philosophy.[23]

This prompts a question: How can a more than 2000 year old philosophical method, developed in relation to a now more or less unanimously rejected ontology of Forms and epistemology of rational insight, not only have survived until this day, but still constitute the standard method of philosophical inquiry? The answer, I will suggest, is that the Platonic picture has been replaced by a largely implicit view on concepts that warrants contemporary analogues of (A_F) and (B_F)—analogues that explain the structural similarity between the Socratic method and contemporary, philosophical analysis.

1.3. Definitional Conceptual Analysis

If we want to understand contemporary conceptual analysis, it is crucial to identify the *object* of analysis. As already noted, contemporary philosophical analysis is concerned with concepts rather than Forms. Unfortunately, however, contemporary analysis has managed to evolve in a way largely disconnected from the psychological study of actual conceptual categorization. An account of the psychological commitments of conceptual analysis, therefore, has to take the form of an inference to the best explanation, identifying the (actual or hypothetical) view of concepts that would make most sense out of conceptual analytic practice.

---

[23] DePaul and Ramsey (1998, pp. vii-viii).

This pursuit will proceed in three phases. In the first, I will characterize conceptual analytical practice as the pursuit of necessary and sufficient application conditions for linguistic terms. I will then identify three desiderata that are prevalent in such practice and use them to extract three substantial hints about the specific structure of the underlying concepts that is assumed by such practice. In the second phase, I will argue that the resulting (proto-)view of concepts is compatible with all three main philosophical theories about concepts—i.e., concepts as abilities, concepts as abstract objects, and concepts as mental representations. This will highlight the sense in which conceptual analytic practice does not commit one to any particular *philosophical* theory of concepts. However, in the third phase, I will argue that conceptual analysis, as traditionally practiced, nevertheless makes most sense against the background of one particular *psychological* theory about concepts, i.e., the so-called Classical Theory of Concepts.

So, as for the first phase, we may note that conceptual analysis in analytic philosophy typically inquires into concepts by way of *defining linguistic terms*. Why would philosophers do this? Why would they suppose that philosophical inquiries into linguistic terms reveal anything philosophically interesting about concepts? My suggestion is that it has been supposed that there is an intimate connection between concepts and the *meaning* of words. The connection, I propose, is the following:

> *Bridge*
> Meaningful words express concepts.

Two comments are necessary. First, *Bridge* does not commit us to saying that there is a one-to-one mapping between concepts and meaningful words. For one thing, there are *ambiguous* words, i.e., words that correspond to several concepts, as well as *synonyms*, i.e., concepts that correspond to several words. For another, there are many concepts that do not have a word (in some cases, not even a string of words) to go with them.

Second, *Bridge* does not imply that the concepts that a person possesses fix everything about what the words she uses *communicate in use*. For example, over and above conveying that someone is armed, "She's got a gun" will communicate different things depending not only on the concepts possessed by the speaker and expressed by the words, but also on the *context* of the utterance, e.g., whether the sentence is uttered at a shooting range or during an armed robbery.

However, none of this serves to undermine the claim we are considering as an explanation of why attending to linguistic terms would help you understand concepts, namely that, when someone uses a meaningful word, that word expresses a (as in: at least one) concept. Furthermore, this claim rests upon a substantial and well-established empirical hypothesis. Hence, Gregory Murphy:

> There is overwhelming empirical evidence for the conceptual basis of word meaning. […] Indeed, I do not know of any phenomenon in the psychology of concepts that could conceivably be found in words that has not been found. If word meanings are not represented in terms of concepts, then they must be represented in terms of something else that just happens to have the exact same properties as concepts. By Occam's razor, I will conclude that word meanings are represented in terms of concepts.[24]

In short, the empirical evidence is such that an impressive amount of the phenomena discovered in the empirical work on human conceptual and, in many cases, *non-verbal* classification, also turn up in the corresponding *linguistic* tasks. For example, people tend to consistently deem some instances of a kind to be more representative or "typical" instances than other. Analogously, sentence planning and word accessibility, as revealed through sentence processing tasks, is sensitive to typicality. Furthermore, human conceptual structure has been shown to exhibit a basic or preferred level, in the sense that, given a neutral

[24] Murphy (2002, pp. 393-394).

setting, people will be more prone to categorize things in terms of some kinds (dog) rather than another (something to be saved in case of fire). Analogously, people are more prone to describe things in terms of words corresponding to such basic kinds, i.e., to describe something as a "cat" rather than a "Siamese," under neutral circumstances.[25]

It should be noted that this story about how meaningful words express concepts remains neutral as to both (*a*) what it is to possess a concept (is it to grasp a Form/Fregean Sense, to instantiate a mental representation, or to have a set of cognitive abilities?) and (*b*) whether meaning, construed thus, serves to determine *reference*. Using a distinction from Kripke, I will, eventually, suggest that, although meaning may not *determine* reference it, nevertheless, serves to *fix* reference, in the particular sense of presenting us with (non-rigid) cognitive pathways to actual, external phenomena.[26] However, if meaning is a function of concepts possessed, the question of reference determining factors will ultimately turn on how we want to understand conceptual *content*.

Since the works of Putnam and Kripke, many philosophers (this one included) have become convinced that the intrinsic, psychological state of the speaker (or what is sometimes referred to as *narrow* content) does not determine reference, at least in the case of natural kind terms and proper names.[27] Rather, reference is, in part, determined by the speaker's environment (yielding so-called *wide* content).[28] Although we will find reason to return to this issue in chapters 2 and 3, the important thing to note for now is that the idea that meaningful words express concepts does not, in itself, commit us to any particular

---

[25] See Murphy (2002, pp. 393-399).

[26] See Kripke (1980, pp. 55 and 57) on fixing reference by way of descriptions and accidental properties, and (1980, p. 96) on the role of such reference fixing in the determination of reference, by way of acts of baptism.

[27] See Putnam (1975b) and Kripke (1980),

[28] See Segal (2000) for a dissenting voice.

view of conceptual content, nor of any particular story about the factors determining reference.

Returning to philosophical analysis, we may note that the favored format for capturing meanings is that of *necessary and sufficient application conditions for linguistic terms*. More specifically, consider the following characterization of what I will refer to as Definitional Conceptual Analysis, or DCA for short:

> *Definitional Conceptual Analysis (DCA)*
> For a given *analysandum* '*F*,' identify a set of characteristics P_1, P_2, … P_n and a function *R* such that *R* takes P_1, P_2, … P_n as arguments in a structure that may involve simple conjunctions as well as other logical connectives and quantifiers, yielding an *analysans* in the form of a definition, citing a necessary and sufficient application condition for the corresponding term "*F*."

In the philosophical literature, three desiderata are commonly invoked in the construction of conceptual analyses, the first of which I would like to characterize thus:

> *Desideratum 1: Simplicity*
> The set of characteristics cited in the analysis (i.e., P_1, P_2, … P_n) should be fairly small and their relation (as modeled by *R*) straightforward, so as to make sure that the resulting analysans is as *simple* as possible.[29]

As noted by William Ramsey, this desideratum is most plausibly motivated in analogy with the explanatory sciences, where it is typically assumed that, if the suggested set of defining characteristics is too complex, disjunctive or convoluted, there is reason to suspect that we simply have not gotten it right yet, and that the definition is the result

[29] Cf. Ramsey (1998, p. 163) and Weatherson (2003, p. 9).

of an *ad hoc* adding of epicycles rather than an accurate characterization of the phenomenon at issue.[30]

A second desideratum that has, to my mind, not been sufficiently acknowledged is the following:

> *Desideratum 2: Exactitude*
> The characteristics cited in the analysis should have *clear-cut* boundaries, so as to ensure that category membership is a straightforward yes-or-no affair.

By invoking this requirement, we are asking for a not just *any* necessary and sufficient conditions but necessary and sufficient conditions that yield an extension that exhibits no fuzzy edges. This requirement is supposed to serve the purpose of ensuring that the extension of the analysans (constructed out of these characteristics) should, in turn, have no fuzzy edges and that it, thereby, provides a clear-cut account of the concept at issue. We will discuss this requirement as well as its motivation in more detail later.

For now, the interesting thing to note is that the most straightforward way to make sense of these two requirements, and the possibility of their joint satisfaction, is that concepts—i.e., the very object of study in conceptual analysis—may be represented via *simple, clear-cut, necessary, and sufficient conditions*. For brevity's sake, let us stipulate that

> *Neatness*
> a *neat* condition is a simple, clear-cut, necessary, and sufficient condition.

As the reader surely notes, this makes for an interesting structural similarity between Platonic and contemporary analysis, with the crucial and already acknowledged difference that Plato is concerned with the essences of Forms while modern analytic philosophy is interested

[30] See Ramsey (1998)

in concepts as expressed by meaningful words. Still, if concepts may be represented by way of neat conditions, it is plausible to assume the following analogue of (A_F):

> (A_C) For every concept '*F*'—or, at least, most philosophically interesting concepts—there is a *definition*, specifying a neat condition for *F*-hood.

As noted already in the previous section, the favored way of not only producing but also evaluating such definitions is by way of categorization intuitions, which is yet another way in which contemporary philosophical methodology is analogous to Plato's. More specifically, it is typically assumed that, if the analysis allows for intuitive counterexamples, it must either (*a*) be accompanied by an explanation of why our intuitions are misguided and, hence, not accurately tracking the concept at issue—in which case I will say that the counterexample is not *genuine*—or (*b*) be considered as not giving an accurate analysis of the analysandum. Hence, our third desideratum:

> *Desideratum 3: Exhaustiveness*
> The definition provided via the condition cited in the analysis should be *exhaustive* in the sense of not admitting any genuine intuitive counterexamples.[31]

Whence this requirement? This brings us to a further similarity between Platonic and contemporary analysis, to the effect that the possession of a concept involves a tacit (albeit potentially incomplete) knowledge of its defining conditions.[32] However, since this knowledge is *tacit*, it is not something that may be straightforwardly produced by any competent user of the corresponding term. Still, in so far as categorization intuitions are the products of the concepts possessed, it is reasonable to assume that there is an elucidatory bridge

---

[31] Cf. Ramsey (1998, p. 163).

[32] Cf. Ramsey (1998, p. 165).

between the categorization intuitions of speakers and the concepts that they possess, such that the former provide substantial hints about the latter. As noted by Goldman:

> It's part of the nature of concepts (in the personal, psychological sense) that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept. If the content of someone's concept *F* implies that *F* does (doesn't) apply to example *x*, then that person is disposed to intuit that *F* applies (doesn't apply) to *x* when the issue is raised in his mind.[33]

Goldman is here concerned merely with concepts "in the personal, psychological sense," by which he means concepts as mental representations—a notion that we will delve deeper into below. It should be noted, however, that *any* reasonable account of concepts has to maintain a correlation between concepts and categorization intuitions. This may be brought out in two steps: First, if concepts are to do any explanatory work, they have to, at the very least, explain why we tend to categorize the world in the ways that we do. In other words, on any reasonable theory of concepts, our categorization dispositions will be non-accidentally correlated with the concepts that we possess. This establishes a link between categorizations and concepts. Second, if forming a categorization intuition is a matter of determining how we would categorize something, our categorization intuitions are non-accidentally correlated with our categorization dispositions. In other words, our categorization intuitions are correlated with our categorization dispositions, which, in turn, are correlated with the concepts that we possess.[34]

Hence, the following analogue of (B_F):

---

[33] Goldman (2007, p. 15).

[34] Cf. Laurence and Margolis (2003, pp. 278-279).

(B$_C$) Probing the categorization intuitions of competent users of a term may serve to elucidate the defining condition that attaches to the corresponding concept.[35, 36]

Just like in the Platonic dialogues, this elucidation plays a *positive* as well as a *negative* methodological role. On the one hand, it provides positive material for candidate analyses, i.e., candidate necessary and sufficient conditions that, under *Simplicity* and *Exactitude*, should be simple and clear-cut. On the other hand, it also serves to evaluate such analyses and, hence, play the potentially negative role of disqualifying them in so far as they either include counter-intuitive instances or fail to include intuitive instances, under the requirement of *Exhaustiveness*.

---

[35] A similar assumption seems to be driving the Chomskyan tradition in linguistics, where the intuitive judgments of speakers are taken as linguistic evidence—on some readings the *only* available evidence—for particular sets of rules and principles of the speaker's language. See Devitt (2006) for a critical discussion and Samuels, Stich, and Tremoulet (1999) for the analogy between the linguistic and the conceptual analytic case.

[36] It might be claimed that there is a methodological problem here in determining when we are dealing with a case of a conflicting intuition of a competent user or an irrelevant intuition of an incompetent user. It seems to me that the only way to separate these is by recourse to the over-all, categorical track-records of users. More specifically, correlations in categorizations between users give rise to (semantic) norms, specifying the correct use of concepts and terms. In so far as a user's track-record of categorization does not deviate to any great extent from this norm, she may be deemed competent and her intuitions taken to carry some weight in elucidations. However, in so far as a user's track-record deviates greatly from the norm, the most plausible explanation is either incompetence or non-standard use—either of which gives us reason to disregard that person's intuitions as far as elucidation goes. Thanks to Åsa Wikforss for calling my attention to this issue.

1.4. On Analysis and Philosophical Theories of Concepts

This concludes the first phase of our inference to the best explanation, prompting the following question for the second: What are the prospects for cashing out (A_C) and (B_C) within the three main philosophical theories of concepts, i.e., concepts as abilities, concepts as abstract objects, and concepts as mental representations?

As for the abilities view, it takes concepts to be nothing but sets of abilities typical to cognitive agents. Hence, *defining* a concept would involve specifying such a set. Under (A_C), this definition would have to take the form of a neat condition delimiting the relevant set of abilities. This particular form might be resisted, and indeed has been resisted by philosophers attracted to the abilities view. Most notably, Ludwig Wittgenstein argued, via several failures to define GAME, that few (if any) concepts have definitions, if understood in the above sense of neat conditions rather than complicated networks of "family resemblances," where category membership need not be a straightforward yes-or-no affair.[37]

More recently, psychologists Eleanor Rosch and Carolyn Mervis have argued that Wittgenstein's claim can be experimentally substantiated by way of so-called Prototype Theory—a theory that we will find reason to look closer at below.[38] For present purposes, however, it is important to note that taking concepts to be abilities does not *in itself* give us any reason to resist (A_C). The same goes for (B_C); there is nothing in the abilities view that hinders one from assuming that the defining conditions attaching to a concept may be elucidated by probing the categorization intuitions of competent users of the corresponding term. In fact, this even seems to be assumed by Wittgenstein, considering the method he uses to rebut purported analyses of GAME: refutation by intuitive counterexamples. Indeed, for such counterexamples to say anything interesting about (our conceptions of) what is *not* a game, they have to track (our conception of) what *is* a game. And although the failure to provide a positive characterization

---

[37] See Wittgenstein (1953).

[38] See Rosch and Mervis (1975).

in terms of neat conditions may lend some credibility to the idea that concept should not be represented thus, it does not discredit (B_C).

Now, turning to concepts as abstract objects. Here the case is somewhat more straightforward, especially in light of the fact that we have already demonstrated the compatibility of (A_F) and (B_F) with one of the main candidates for concepts as abstract objects, i.e., Platonic Forms. Hence, assuming that concepts just *are* Forms, we have also shown the latter to be compatible with (A_C) and (B_C), *mutatis mutandis*. It does not take much effort to extend this case to the other main candidate: Fregean senses. On this view, to possess a concept is to grasp a sense.[39] Such senses play the dual theoretical role of (*a*) determining the referent and (*b*) accounting for the mode in which the referent is presented and, thus, for the possibility of cognitive differences between co-referring expressions (such as "seven plus four" and "the square root of one hundred and twenty one"). What would it mean to define the sense of an expression? It would mean to specify the description that a (particular or generic) subject grasps when accessing a referent. In other words, to possess a concept is to grasp the description in question. If so, however, it seems fairly reasonable to assume that the categorization intuitions of competent users of particular terms provide substantial information about that description. Hence, assumption (B_C). For (A_C) to be warranted, however, this description needs to take the form of a definition in terms of a neat condition. Just like in the above, this is, clearly, something that one might have independent doubts about, but it is not ruled out by the mere fact that concepts are senses—which is all that matters for present purposes.

This brings us to our final philosophical theory of concepts: concepts as mental representations.[40] On this view, concepts are men-

---

[39] To anyone familiar with contemporary terminology, Frege's taxonomy is slightly confusing here. To Frege, a sense ("*Sinn*") is not the same thing as a concept ("*Begriff*"), which is the referent ("*Bedeutung*") of a predicate.

[40] See, e.g., Fodor (2003) and Pinker (2007; 1994). It should be noted that understanding concepts in terms of mental representations in no way com-

tal tokens of particular types, playing a direct causal-functional role in the mental life of subjects. To have a concept is to token such a type and for two subjects to share a concept is simply to token the same type.[41] This way of thinking about concepts has been particularly influential in the interface between philosophy, cognitive science, and empirical psychology. It is important to note, however, that the philosophical relevance of empirical research on human categorization is in no way contingent upon concepts being mental representations, rather than abilities, Platonic Forms or Fregean senses. As noted in §1.3 above, the very *Bridge* tying together the inquiry into application conditions for words with the structure of concepts turns on the substantial, empirical hypothesis that meaningful words express concepts. However, agreeing to this does not commit one to a particular *ontological* theory of concepts. Platonists may still claim that concepts are Forms and using a meaningful word is a question of grasping such Forms, just like the Fregean may claim that concepts are senses and that using a meaningful word is a question of grasping such senses. Similarly, an adherent of the abilities view may hold that concepts are nothing but sets of cognitive abilities, e.g., pertaining to discrimination and inference, and that the meanings of words are satisfactorily explained with reference to such abilities. Finally, psychologists might prefer to think about concepts in terms of mental representations. I happen to share this preference and will, henceforth, refer to mental representations by the term "concept," unless otherwise is stated.

However, this does not take away from the fact that what is ultimately being studied by psychologists interested in concepts is human categorization, i.e., facts pertaining to the ways in which people

---

mits one to the idea that concepts can be analyzed. See e.g., Fodor (1981). See also Pinker (2007, pp. 92-102) for a critical discussion.

[41] As noted by Putnam (1975b, p. 222), Frege's reluctance to identify concepts with psychological states seems to be a result, at least in part, of overlooking this latter possibility, with the consequence that "Frege's argument against psychologism is only an argument against identifying concepts with mental particulars, not with mental entities in general."

categorize the world. And unless one wishes to sever the (explanatorily quite potent) connection between such categorization and the concepts that, supposedly, explain the structure of and patterns of our categorizations, one will have to agree to that

> *Constraint*
> the classificatory structures that arise out of concept use offer substantial constraints on the correct theory of concepts, in such a way that, given certain patterns in classificatory structures, we may rule out certain theories as incorrect.

As seen above, DCAs are typically constructed in the form of definitions, citing neat application conditions for terms, supposedly capturing the meanings of concepts. However, as we noted in passing above, it has been suggested that the structure of human concepts—i.e., the very object of conceptual analytic study—as revealed by the classificatory structures that arise out of concept use, are not satisfactorily captured by neat conditions. More specifically, we will now see that the classificatory structures unveiled by psychologists studying human categorization display properties incompatible with the idea that concepts exhibit the structure of neat conditions, which, given *Constraint*, seems to imply that (A_C) is not scientifically warranted.

1.5. The Classical Theory of Concepts

There was indeed a time when psychologists assumed that concepts were best captured by neat conditions—a view that is usually referred to as *the Classical Theory of Concepts*. This view may be summed up as follows:

> *The Classical Theory of Concepts*
> Concepts are best described in terms of definitions, providing necessary and sufficient conditions for category membership in such a way that
> (1) there are no distinctions between category members; and

(2) for every category, every object is either in or not in that category.[42]

However, researchers soon found reason to doubt this classical picture, initially through the aforementioned work of Rosch and Mervis.[43] First of all, it turned out to be extremely hard to find *any* concept, the meaning of which could be summed up through a neat condition—a point that, as we have seen, was made already by Theaetetus in his discussion with Socrates and later by Wittgenstein. More importantly, however, and as for the particular characteristics called for by (1), it was found that the categorizations we, in fact, make reveal a taxonomy where members of a category form a *continuum*, and some members (often quite consistently) are judged as better examples of the category than others—a phenomenon that has come to be referred to as *the typicality effect*. In other words, while some entities (such as trucks and tablecloths) clearly do not qualify as birds, some entities within the category of birds are clearer examples than others.

For example, even in cases when the relevant traits we, supposedly, would include in a list of necessary and sufficient characteristics are equally salient, robins are considered more typical instances of birds than eagles, which is revealed in reaction time experiments on subjects' tendencies to pair specimens with particular kinds.[44] Furthermore, and as for (2), it was also found that, rather than categorizing objects against the background of neat conditions, category membership is a question of similarity to typical instances along different dimensions, where some features are more "important"—i.e., get assigned a greater *weight*—than others. And, in many cases where the similarity between the typical instances and a judged item gets lower,

---

[42] This formulation is borrowed from Murphy (2002, p. 15), with some slight re-formulations to fit the taxonomy of this study.

[43] See Rosch and Mervis (1975).

[44] See, e.g., Rips, Shoben, and Smith (1973). See also Rosch (1977).

there is no clear answer to the question whether the item is or is not in the category.[45]

Clearly, the rejection of the Classical Theory presents a problem for DCA. More specifically, while not directly refuting the idea that concepts may be captured via necessary and sufficient conditions, the aforementioned empirical results do serve to challenge the assumption that concepts should be construed along the lines of (1) and (2) above, i.e., as yielding categories where no distinctions are made between category members and every category has clear-cut borders that delimit it from every other category.[46] As such, the rejection of the Classical Theory directly discredits *Exactitude*—i.e., the "clear-cut" aspect of neatness—as a desideratum for analysis and for the following two reasons: First, given that human categorization, *contra* (1), reveals a continuum structure in light of the typicality effect, such that some members are deemed to be "better" instances than others, it is plausible to assume that some "bad" members will not be all that different from some non-members, and that there, hence, seldom will be any sharp boundaries between different categories. This very assumption is made further plausible by the evidence that, *contra* (2), suggests that categories are generated in reference not to clear-cut characteristics but rather to similarity to typical instances. Hence, the kind of characteristics called for by *Exactitude* simply does not seem to be of the right kind, if we are interested in capturing human concepts.

As noted by Ramsey, we may even find some support for a rejection of the Classical Theory by looking at actual philosophical dialectics in relation to DCA conducted under *Exactitude*—at least if combined with *Exhaustiveness*.[47] Against the background of these two desiderata, an instance of DCA may be refuted in either of two ways. On the one hand, it may be refuted through the identification of a genuine intuitive instance that does not possess all the properties cited in the definition, which would show that the characteristics are not

---

[45] See, e.g., Murphy (2002, pp. 30-31).

[46] Cf. Margolis and Laurence (1999, p. 24).

[47] See Ramsey (1998).

necessary. However, granted what psychological research tells us about the structure of concepts, it is to be expected that such instances are not too hard to come across. All we have to do is identify a situation in which one or more of the proposed definitional features (say, ability to fly) are lacking but the summed weight of the ones that are present (feathers, beaks, wings, vertebrate, egg-laying, etc.) is sufficient to, nevertheless, yield an inclusive verdict.

The other way in which analyses may be refuted is through the identification of an instance that has all the features cited in the definition but that, nevertheless, does *not* intuitively fall within the category in question, which would show that the features in question are not *sufficient*. Again, such instances might not be too hard to find, given that any of the two following claims holds: (*a*) Feature weight assignments are context-dependent, in which case it is possible frame counterexamples in terms of contexts that tip the scale towards negative characterization, and, thereby, enable feature distributions with a summed weight that normally yield an inclusive verdict (a subject is an unmarried man, hence, he is a bachelor) to fail to do so because of contextual factors (the subject happens to be the Head of the Catholic Church); or (*b*) features might cancel each other out when figuring in non-standard scenarios, so that a set of features that would normally lead one to judge an instance as a member of a particular category (*x* is a device with a seat-cushion and a back-rest and is designed to sit in; hence, *x* is a piece of furniture) does not get categorized thus, since it has a feature that, in particular circumstances (say, if installed in a car), is more common to another category (a car seat).[48]

In conclusion: Given that concepts are represented in a way that best accommodates the phenomena uncovered by psychologists critical to the (today almost unanimously rejected) Classical Theory, it is to be expected that analytical philosophy, if practiced along the lines of DCA under the requirement of *Exactitude* and *Exhaustiveness*, generates an abundance of counterexamples and refuted theories. And

[48] See Ramsey (1998, p. 171-2) and Smith and Medin (1981).

everyone familiar with the analytical philosophical tradition would be hard pressed to deny that this is what we, in fact, have seen.

1.6. Whence the Desire for Exactitude?

One question remains, however: If the Classical Theory of concepts is so untenable, how come philosophers have been so attracted to the idea of characterizing concepts in terms of clear-cut definitions? One reason is that it presents a unified theory of several heavily researched phenomena, the three most important being concept acquisition, concept categorization, and the determination of reference.[49] If possessing a concept means possessing a neat condition, we may say that learning a concept just involves *acquiring* such a condition, that categorizing just involves *applying* such a condition, and being the referent simply involves *satisfying* such a condition.

Undoubtedly, this is a very attractive story and it is not made less attractive by the fact that it is able to subsume the explanations of three very important phenomena under one simple and unified theory. In light of this, it also makes sense that it was to be rejected first when all three components had been shown implausible, with the 1970's marking the beginning of the downfall. We have already reviewed some of the results regarding typicality, discrediting this picture in relation to categorization. As it happened, the same phenomena were found also in the categorizations made by children, further discrediting the idea that we acquire concepts via clear-cut definitions.[50] As a final blow, the works of Putnam and Kripke provided compelling reason to believe that—at least as far as proper names and natural kind concepts go—we are hardly ever in possession of conditions specific enough to determine the referents.[51]

Still, this does not necessarily explain why philosophers concerned with conceptual analysis have focused on *clear-cut* definitions.

---

[49] Cf. Margolis and Laurence (1999, p. 10).

[50] See Murphy (2002, pp. 318-319 and 336-340) for a recent overview.

[51] See Putnam (1975b) and Kripke (1980).

This brings us to another important and historically influential motif: the idea that the products of conceptual analysis should, ideally, be incorporable into axiomatic systems, i.e., systems constructed out of a finite (and usually small) set of axioms, a (potentially infinite) set of theorems, and a set of inference rules specifying how to infer the latter from the former. This idea can be found already in the writings of Plato but figures in a slightly more worked out form in the works of his student Aristotle. Just as his tutor, Aristotle considered deductive science to be the most noble and important form of science and argued that the proper logical structure of such science is axiomatic. More specifically, the sentences of such sciences should either correspond to axioms or be derivable from them by way of inference rules.

More importantly for our purposes, an analogous requirement is put on the component terms of such deductive sciences, which are to be introduced either without any definitions, i.e., as basic terms, or to be defined on the basis of such basic terms. One very intuitive way to conceptualize such axiomatic systems is in relation to set theory. The sets of set theory are abstract objects that serve to define other concepts via a membership relation, unions, and the intersections of sets. Traditionally, membership in sets has been construed in such a way that (*a*) no distinction is made between different members of the same set, with the consequence that set-membership is a pure yes-no question, while (*b*) a very clear distinction holds between members and non-members of a set, to the effect that the borders that delimit sets are completely clear-cut.[52]

If philosophy is to be conducted in accordance with this axiomatic ideal, concepts—i.e., the targets of philosophical analysis—must be tailored to fit this picture. And, undoubtedly, construing concepts as determinate sets is not without its advantages. For one thing, it provides a very useful framework for specifying what it is for two concepts to be identical versus contrary (the sets completely coincide versus do not intersect) as well as to be similar and/or dif-

[52] More recent attempts to re-conceptualize sets as *fuzzy* sets, where set membership rather is a matter of probability or degree, deny the first assumption.

ferent (the sets intersect to this-or-that degree). Furthermore, and as for properties in an axiomatic context, a set-theoretic framework also yields clear criteria for what warrants *inferring* one concept from another: the set corresponding to the inferred concept is either a superset of the set corresponding to the concept from which it is inferred (as when inferring X IS BLUE from X IS LIGHT BLUE), or the sets completely coincide (as when inferring X IS CRIMSON from X IS OF A RICH DEEP RED COLOR INCLINING TO PURPLE). In other words, determinate sets, clearly, provide a powerful tool in the construction of axiomatic systems of concept.

As we have seen, however, the classificatory structures that arise out of actual, human classification do not lend themselves to a characterization in terms of such determinate sets, which has direct implications for philosophical methodology. Or to put the point more bluntly, as William Lycan has done recently in a retrospective piece on the Gettier discussion:

> It is well to remind ourselves that no effort of analytical philosophy to provide strictly necessary and sufficient conditions for a philosophically interesting concept has ever succeeded. And there should be a lesson in that.[53]

So, what is the lesson? Clearly, something needs to be learned, but it is too early to be pessimistic about the project of analyzing concepts as such. The next chapter considers two rectifications of traditional methodology, the first one in terms of prototypes and the second one in terms of reflective equilibria. This will serve to highlight yet another problem for conceptual analysis—does an essentially armchair method provide the best methodology for understanding concepts, in light of the more rigorous methods of empirical psychology?—and, ultimately, lead us to the more fundamental question: Why should we analyze concepts in the first place?

---

[53] Lycan (2006, p. 150).

1.7. Conclusion

There is a striking continuity between the Socratic method of Plato's dialogues and that of contemporary philosophy, such that analysis is construed as the pursuit of *definitions* by way of *intuitions*. I have argued that one way to understand this continuity is with reference to how the Platonic idea that philosophical insight corresponds to insight into the *essences* of eternal Forms has been replaced with the contemporary idea that the proper objects of philosophical investigation are the *meanings* of terms as represented by the structures of concepts. Despite this shift in target, however, one central idea has been preserved, namely that the targets of analysis are best characterized in terms of simple, clear-cut, necessary, and sufficient conditions. However, when turning to our best psychological evidence regarding the structure of our concepts, we found little support for the idea that such conditions provide the best format for representing concepts. In the next chapter, we will consider two attempts to rectify conceptual analysis, the first one in terms of prototypes and the second one in terms of reflective equilibria.

Chapter 2.
Prototypes and Reflective Equilibria

In a sense, the lesson of the previous chapter is perfectly straightforward: Any philosopher concerned with the analysis of concepts—be they epistemic or not—has to take the empirical work on actual human categorizations seriously. This also highlights the sense in which the problem about neat conditions constitutes an *internal* objection to conceptual analysis, in the sense that it does not discredit the project of analyzing concepts *as such*, but merely throws doubt on a particular way of analyzing.

For this reason, the problem seems perfectly solvable; all we need to do is find an empirically more warranted way to characterize concepts. In the present chapter, we will first consider a solution in terms of what I will call Prototypical Conceptual Analysis, working with concepts as prototypes rather than neat conditions. This will not only highlight questions regarding whether the armchair provides a satisfactory methodological vantage point for the understanding of concept, given the more rigorous methods of empirical psychology, but also certain issues regarding why we should analyze concepts in the first place—issues that will, ultimately, drive us to question *Exhaustiveness* as a plausible desideratum for epistemological analysis.

This will bring us to the second solution in terms of the idea that the goal of analysis is not to provide exhaustive accounts of our concepts, but rather to construct theories that put our most central intuitions in a reflective equilibrium with our general principles or theories about the world. It is argued that, construed thus, this idea, at best, comes out to a call for honest, intellectual inquiry, grounded in our best empirical theories. Needless to say, such inquiry should play an important role in philosophical inquiry, epistemology being no exception. However, the challenge lies in providing methodology that specifies exactly what weights to assign to our current concepts and norms in philosophical theorizing, exactly at what stage empirical inquiry enters, and exactly what is the mark of a good analysis—challenges that will be taken up in chapters 3 and 4.

2.1. PROTOTYPICAL CONCEPTUAL ANALYSIS

While being fairly conclusive as for the refutation of the Classical Theory, the psychological research is, unfortunately, suggestive at best when it comes to exactly what theory should replace it. Nevertheless, on one popular view, concepts are represented as having the structure of *prototypes*, understood as abstracted sets of typical features. As was hinted in the previous chapter, categorization, on this view, is a similarity comparison process. More specifically, it is a function of the number and weight of prototype features possessed by the categorized item, where the weight signifies the importance of the feature in question, mirroring the assumption that some features are more important than others. For example, although both the property of doing harm and being made of metal probably figures the prototype for WEAPON, the presence of the former is, clearly, more important than the latter.[1]

So, let us evaluate the prospects for doing conceptual analysis in terms of prototypes rather than neat conditions, by way of what I

[1] See Murphy (2002, pp. 42-43).

will refer to as *Prototypical Conceptual Analysis*, or PCA for short. Such an analysis, I suggest, would take the following form:

> *Prototypical Conceptual Analysis (PCA)*
> For any '*F*,' identify a prototype set *Q* and a threshold value *T*, such that *Q* contains a set of prototypical features, assigned appropriate weights in such a way that the weighted sum predicts positive categorization for '*F*' when, and only when, it exceeds *T*.[2]

Construed thus, it is important to note is that PCA is *not* working under assumption (A_C) but rather under the following analogue:

> (A_P) For any concept '*F*,' there is a set of prototypical features assigned weights that, together with a threshold value, predicts categorization.

Let us look closer at this assumption in relation to the three desiderata set up earlier. It was noted above that characterizing concepts in terms of neat conditions failed to do justice to the dual fact that (*a*) instances often form a continuum where some instances are quite consistently considered "better" instances than others, and that (*b*) there is not always a clear answer to the question whether a particular item is or is not an instance. Both of these facts may be accounted for when concepts are construed as prototypes. Given that items may have more or less of the weighted features in *Q*, items that have enough features to exceed *T* will form a continuum in accordance with (*a*). Furthermore, given that *T* may be construed either as an absolute value, or a value satisfied in so far as it is approximated, the present model can not only handle situations calling for a specific cut-off value, but also situations with vague concept boundaries, in accordance with (*b*). Hence, PCA rejects *Exactitude*, in allowing for the

[2] Goldman and Pust (1998, pp. 193-194) seem to have something like this in mind. See also Goldman (2007, p. 23).

possibility of intra-concept distinctions and non-clear-cut boundaries rather than neat conditions—a focus well motivated by the scientific findings reviewed in relation to DCA above.

Furthermore, if we hold on to *Exhaustiveness*, PCA latches on to the features and weights assumed by (A_P) and yields predictions of categorization judgments. As such, PCA admits of no counterexamples. However, as already the above characterization makes clear, PCA would under *Exhaustiveness*, most likely, result in utterly complicated analyses, and, thereby, fail to satisfy *Simplicity*. Still, if one looks at actual philosophical practice, PCA is no different from DCA here, which, in a context where it is to be expected that counterexamples are cheap (as was argued in the previous chapter), has generated notoriously complicated analyses of seemingly ordinary notions. Unlike DCA, however, PCA would not generate complicated analyses due to an inaccurate theory of concepts. Rather, PCA would yield complicated analyses because concepts and categorizations are, as a matter of fact, constituted and governed by quite complicated mechanisms.

2.2. Conceptual Analysis as an Empirical Task

This observation also serves to highlight the point that conceptual analysis may be construed as an explicitly *empirical* task—i.e., as relying on (current) sense experiential input—which is in stark contrast to how it has been conducted by analytic and armchair prone philosophers. Considering the common opinion that conceptual analysis is *a priori*, it will serve us well to elaborate on the distinction between empirical and armchair (i.e., non-empirical) inquiry in relation to the distinction between *a priori* and *a posteriori* warrant. First, what is *a priori* and *a posteriori* warrant?

As I will use the terms, *a priori* warrant is warrant that appeals to pure thought or reason alone, and *a posteriori* warrant is simply warrant that is not *a priori*.[3] In explicating the relevant sense of "appeals to" it has become customary to distinguish between *evidential*

[3] Cf. BonJour (1998, p. 11).

and (merely) *enabling* factors. Hence, while visual perception may be evidentially relevant to my belief that the sky is blue, it plays a merely enabling role in my belief that everything that is blue is colored; without visual experience, I would (arguably) not have been able to form the concepts BLUE and COLORED, nor raise the question whether all blue things were colored. However, as soon as I have acquired these concepts, I will be able to conclude that everything that is blue is colored by recourse to pure thought or reason alone, without any evidential input from perceptual experience or the like. Against the background of this distinction, we may reformulate our characterization of the *a priori* and *a posteriori* as follows: *a priori* warrant is warrant that appeals to pure thought or reason alone as far as *evidential* factors go, and *a posteriori* warrant is simply warrant that is not *a priori*.

Now, do the two distinctions—i.e., empirical and non-empirical method, on the one hand, and *a posteriori* and *a priori* warrant, on the other—coincide? No, they do not. It should be clear that *a priori* warrant never involves an empirical method, but it hardly follows that *a posteriori* warrant never involves a non-empirical method. Some paradigm examples of an empirical method do, indeed, proceed by way of *a posteriori* sources of warrant—knowledge through perceptual observation being one of them—but this does not take away from the fact that some warrant is neither *a priori* nor flows in any straightforward way from sense experience. Take introspection and memory, for example, neither of which can be plausibly said to give rise to *a priori* warrant. However, since they do not rely on (current) input provided by our five senses in any obvious way, they would still qualify as non-empirical sources of warrant under the above characterizations.

The following table sums up and exemplifies the suggested relation between the distinction between the *a posteriori* and *a priori*, on the one hand, and that between an empirical method and a non-empirical method, on the other:

	A priori warrant	*A posteriori* warrant
Empirical method	-	Perceptual observation.
Non-empirical method	Mathematics and logics.	Memory and introspection.

As suggested by Goldman, this way of construing the *a priori*/*a posteriori* distinction has the advantage that it enables us to formulate two distinct approaches to conceptual analysis and the use of categorization intuitions. First, take *the third-person approach* to conceptual investigation:

> The experimenter presents a subject with two verbal stimuli: a description of an example and an instruction to classify the example as either an instance or a non-instance of a specified concept or predicate. The subject then makes a verbal response to these stimuli, which is taken to express and application intuition. This intuition is taken as a datum—analogous to a meter reading—for use in testing hypotheses about the content of the concept in the subject's head.[4]

As Goldman notes, the resulting evidence is "distinctly observational, and hence empirical." More specifically, this is an example of conceptual analysis construed in empirical (i.e., sense experiental), *a posteriori* (i.e., non-*a priori*) terms. However, it does not look much like the way in which conceptual analysis has traditionally been conducted. This brings us to *the first-person approach* to conceptual analysis, where one is primarily consulting one's own categorization intuitions.[5] Now, construed thus, conceptual analysis, clearly, does not involve making any perceptual observations. However, on the above definitions, it does

---

[4] Goldman (2007, p. 20).

[5] Note that calling this the *first-person* approach is only meant to imply that we are dealing with a person consulting her *own* intuition—not that she necessarily has a privileged access to the matters at hand.

not, thereby, follow that the resulting warrant is *a priori*. Hence, Goldman:

> Since some sources of warrant are neither perceptual nor a priori, application intuition might be another such source. Indeed, the process of generating classification intuitions has more in common with memory retrieval than with purely intellectual thought or ratiocination, the core of the a priori. The generation of classification intuitions involves the accessing of a cognitive structure that somehow encodes a representation of a category. Of the various sources mentioned above, this most resembles memory, which is the accessing of a cognitive structure that somehow encodes a representation of a past episode.[6]

In other words, conceptual analysis from a first-person approach is best described as an instance of a non-empirical (i.e., non-sense experiential), *a posteriori* (non-*a priori*) investigation.

Now, let us return to PCA and pose the question: What is the most promising approach to conceptual analysis conducted in terms of prototypes, the first-person or the third-person approach? In answering this question, we need to keep in mind that we, on PCA, not only need to determine the prototypical features of the concept in question, but also the weight of each feature, not to mention any contextual factors that might influence the actual assignment of such weights. And make no mistake: This is all as it should be, if we want fully exhaustive characterizations of our concepts. And, as such, the empirical study has, as already noted, been conducted with impressive assiduity within psychology.[7]

More specifically, consider the following reformulated analogue of (B_C):

---

[6] Goldman (2007, p. 20).

[7] See Smith and Medin (1981) and Murphy (2002) for two good overviews of the psychological study of concepts.

(B_P) Probing the categorization intuitions of competent users of a term may, under very special circumstances, serve to elucidate the prototypical features and weights that attach to the corresponding concept.

This brings out another internal objection to conceptual analysis: If we are to hold on to the ambition of exhaustively characterizing concepts, conceptual analysis has to go beyond mere armchair exercises in favor of hands-on empirical investigation, since mere armchair reflection is a less reliable guide to our concepts than empirical science. There are (at least) three reasons for this. First, the methods of empirical science is better suited for coping with performance errors on part of the intuiting subject, i.e., with situations in which our intuitions are off the mark due to human limitations in attention span, computational capacity, and the like. Second, when compared to the experimental condition of the single, intuiting subject, empirical science has superior resources for not only collecting but also handling large sets of data. Third, due to greater methodological rigor, as well as increasingly sophisticated technical apparatus, empirical science is more likely to provide a more exact picture of our concepts. After all, it was not until the dawn of scientific, psychological study of concept that such subtle but important phenomena as typicality were discovered.

Note that this is not a call for giving up on intuitions altogether. This is a point about methodology—not skepticism. As such, it does not aim to reject the idea that we might very well have an intuitive grasp of the contents of our concepts along the lines of the first-person approach. Indeed, any inquiry into our concepts will rely on introspective reports regarding such content. It *is* to deny, however, that the armchair provides the most methodologically sound location for the analysis of concept, given the more rigorous methods of empirical psychology, that typically not only proceeds by way of a more substantial body of data (as in: data that is more substantial than what is provided by the introspective reports of a single philosopher

and her colleagues) but also takes seriously the methodological problems that face psychological research in general and the use of introspective reports in particular. Hence, the formulation "Under very special circumstances" in (B_P) above.

Here, it should also be noticed that, since the inception of prototype theory in the 1970's, much attention has been paid to a question that many defenders of the theory has failed to address: What are the mechanisms determining what features get included in the prototype set and the weights that these features get assigned? This is a question best answered in the context of conceptual development. As it turns out, prior knowledge plays an important role in the construction and evolution of a prototype set in the sense of providing substantial constrains not only on what features are allowed in the set but also on the weights that are assigned to these features.

For example, in the case of artifactual kinds, little attention is typically paid to superficial properties with no relation to design and purpose, quite independently of how often these properties occur in instances.[8] Similarly, people forming concepts of animals typically pay no attention to such properties as the age or sex of the instances. In fact, in the case of natural kinds, people tend to be fairly realist in their categorization habits, treating the properties used to pick out instances not so much as defining features but something closer to symptoms of an underlying essence—an approach that gives rise to increasingly sophisticated feature sets that are better predicted by factors pertaining to prior knowledge about the kind of features that are important (e.g., for inductive inference) than by simply keeping count of how often particular features appear in category members.[9] Consequently, prototype theories, at least as traditionally formulated, fail to predict the subtleties of conceptual development.

These are all very interesting issues and any serious attempt to defend a prototype approach to conceptual structure would have to involve a thorough treatment of them all. Indeed, this is exactly what

---

[8] See, e.g., Murphy (1993, 2002).

[9] See, e.g., Keil (1989).

we find within the psychological literature on concepts.[10] However, we need to remember that we are here interested in conceptual analysis as it figures in philosophy generally and in epistemology in particular. It was noted above that conceptual analysis has traditionally been conducted under the assumption that analysis should provide *exhaustive* accounts of our concepts, i.e., analyses that reveal every conceptual nook and cranny. Clearly, this is not an unreasonable desideratum within the branch of psychology concerned with the study of concepts. Indeed, nothing short of exhaustive accounts would reveal exactly those subtleties that we need to understand in order to get a better grip on the exact structures of concepts and the way in which they develop. However, is *Exhaustiveness* an important desideratum in the *philosophical* analysis of concepts?

To the extent that it is, the above considerations would force philosophers concerned with conceptual analysis to not only take into account the problems discussed in chapter 1 in relation to neat conditions and the fall of the classical theory, but also face up to the challenges involved in providing a positive theory of concepts that is more in keeping with the available empirical evidence. However, keeping in mind the possibility of a difference in desiderata between psychologists concerned with unveiling the mental mechanisms governing our categorizations, and the epistemologist concerned with describing and aiding the epistemic inquirer, it is time to ask ourselves: Why should epistemologists be interested in exhaustive accounts of our epistemic concepts in the first place?

2.3. Norms, Goals, and Epistemic Architectures

If we take a look at the literature engaged in the analysis of epistemic concepts, we note that epistemologists are not interested in analyzing just any concepts. The ones of interest are the particularly *normative*

[10] See Murphy (2002, pp. 183-190) for a recent overview and discussion, including the extent to which the problems in question extend to so-called exemplar theory—a close cousin to prototype theory.

concepts. The easiest way to understand how such concepts work is by contrasting them with *non*-normative concepts such as TABLE and DOG. By employing such concepts—indeed, by employing *any* (non-syncategorematic) concept—we are, clearly, concerned with *categorization*, i.e., with grouping particular entities under general terms. As such, concepts like TABLE and DOG may be put to use within an explicitly normative *context* pertaining to (the norms of) proper word use, as in "You shouldn't call Fido a cat—he's clearly a dog."

This, however, does not make the statement "Dogs are thus-and-so" normative, which we may see if we contrast them with such concepts as KNOWLEDGE, JUSTIFICATION, and RATIONALITY. The latter are all concepts by which we may evaluate fellow epistemic inquirers—that is, categorize their conduct as being or not being an instance of knowledge, justification, rationality, etc.—but where we, *merely by virtue of such a categorization*, also make an explicitly normative judgment about the extent to which what they are doing is good from an epistemic point of view. More specifically: Unlike "that is a dog," the statement "Paul is justified" has normative implications since the latter, but not the former, is not only associated with a set of semantic norms regarding proper word use, but also a set of particularly *epistemic* norms according to which you *should* be justified, since being justified is, supposedly, conducive to certain epistemic goals.

So, what are epistemic norms, then? Here, epistemic norms will be understood as hypothetical imperatives of the following form:

Given epistemic goal *G*, every element in *A* should (be) *F*,[11]

where *A* is a set of epistemic subjects, possibly containing just one member, and '*F*' any arbitrary epistemic concept. In other words, the normative framework I am working with here is explicitly *instrumental-*

[11] An alternative formulation would be "Given goal *G*, every element in *A* should (be) *F*, under circumstances *C*." However, as long as *C* is incorporable into the conceptual component '*F*,' characterizing norms and principles in terms of the more general form will not result in any lack of precision.

ist, in that it takes epistemic normativity to be a question of the extent to which something constitutes or is a means to an epistemic goal. For the sake of brevity, I will refer to anything constitutive of or conducive to an epistemic goal—i.e., anything that may be properly plugged in for "*F*" in the above formulation—as an *epistemic desideratum*.[12]

However, as will become more obvious as we go along, it serves us well to distinguish between two kinds of normativity. I will refer to the first kind as *general normativity*. General normativity simply concerns what is good in the instrumental sense of being constitutive of or conducive to a goal. As such, general normativity does not take into account what an agent may or may not bring about voluntarily. For example, as epistemic agents, we strive to attain certain epistemic goals, one central and important being true belief. (This goal will be qualified and discussed at length in chapter 4.) Given this goal, it is, clearly, something good to believe truly, just like it might be good to digest properly, even though our digestive apparatus is not within our voluntary control. For this reason, the following norms are endowed with general epistemic normativity, where the italicized phrases designate desiderata:

> *S* should *believe truly*;
> *S* should *believe by way of reliable belief-forming processes*.

Now, clearly, believing truly and reliably are good things, even though we cannot just believe truly or reliably at will. At the same time, there are several things we can do to put ourselves in a position of believing truly or reliably. This motivates the introduction of a more fine-grained notion of normativity: *action-guiding normativity*. Action-guiding normativity takes into account not only what may be more or less good in relation to a specified set of goals, but also what agents may

---

[12] Note that the notion of an epistemic desideratum that I am working with here differs from Alston's (2005), who uses the same term to specifically designate desirable features of belief.

or may not do to put themselves in a position of meeting those goals. For example, given that true belief is a goal of *S*'s, the following are plausible candidates for norms endowed with action-guiding normativity, where the italicized phrases designate desiderata:

> *S* should *scrutinize her grounds for belief by way of introspection*;
> *S* should *attend more to the advice of experts than novices*;
> *S* should *take care to review her evidence before passing judgment*;
> *S* should *not engage in excessive guesswork or wishful thinking.*

The particular merits of the aforementioned norms—i.e., whether or not they actually are more or less appropriately related to our epistemic goals—will be discussed at length in the second part of this study. For now, it is important to note that it does not follow from the fact that we *take* certain activities to be plausible means to attain our goals, that those activities *do*, in fact, constitute plausible means to attain our goals. For example, in ancient Greece, the prophesies of the Pythia, the priestess presiding over the Oracle of Apollo at Delphi, were taken to provide substantial information about the future, and kings regularly consulted her in highly significant matters, such as whether or not to go to war. In other words, many people in Ancient Greece would most likely subscribe to something like the following norm:

> *S* should *give high credence to the predictions contained in the prophecies of the Pythia.*

More than this, people did not just subscribe to this norm for any reason; they subscribed to it because they believed that it enabled them to increase their chances of having true beliefs about the future. Today, however, we would take them to be mistaken in believing this, reject this norm in favor of more modern norms of predictions, and perhaps explain the Pythia's vivid "prophecies" with reference to the hallucinogenic vapors rising from the Castalian Spring that surrounded her. That is, without denying that the above norm might

have played an ever so important role in Ancient matters of prediction, we would want to argue that people in Ancient Greece were mistaken in taking it to be conducive to their (and our) epistemic goal of attaining true beliefs about the future.

Furthermore, as will be discussed further in chapter 4, it should be noted that *S* having certain epistemic goals does *not* imply that these goals may not be overruled by non-epistemic considerations in more naturalistic settings. In other words, that it is a goal of *S*'s to have true beliefs does not imply that *S* should have true beliefs *all things considered*. The presence of such a goal is perfectly compatible with that *S*, in many situations, should *not* have true belief, due to certain non-epistemic considerations pertaining to, say, (decision-making) speed, cost-effectiveness, etc.

By way of taxonomy, I will in the following refer to both general and action-guiding normativity when talking in terms of norms and normativity, unless otherwise is specified. Furthermore, I will refer to conglomerates of concepts, norms, and goals as *epistemic architectures* (at this stage, without passing any judgment on their epistemic merits) designating both fairly simple sets as well as sets involving a rich variety of concepts, norms, and goals in increasingly complex constellations, representing actual as well as merely possible (and more or less promising) frameworks for epistemic inquiry.[13]

---

[13] My notion of epistemic architectures bears some similarity to Goldman's (1992b) *epistemic folkways*. However, unlike Goldman, I will (*a*) not assume that there is enough conceptual homogeneity to warrant talk about *our* epistemic folkways, for reasons that will be discussed in chapter 3, and (*b*) not only include concepts and principles but also epistemic *goals*. As we will see in chapter 4, it seems more plausible to assume that there is less of a variation in the epistemic goals than in the epistemic concepts of different architectures or epistemic (folk)ways. However, see Goldman (2001, p. 477), where he seems to concede the latter point.

2.4. EPISTEMIC ROMANTICISM

Taking such epistemic architectures as an important epistemological object of study, we may now identify a candidate rationale for why we should analyze epistemic concepts:

> (SYNC) Our epistemic concepts and norms are in full sync with our epistemic goals, in the sense that, by adhering to the norms in which our epistemic concepts figure, we are presented with the optimal way of reaching our epistemic goals.

If (SYNC) holds, we have reason to embrace what Weinberg, Nichols, and Stich have referred to as *Epistemic Romanticism*, i.e., the thesis that the *correct* epistemic norms are, somehow, already implanted within us and (assuming introspective access) discoverable with the proper process of self-exploration.[14] Consequently, the *descriptive* task of exhaustively analyzing our epistemic concepts (and, indirectly, the norms in which they figure) may be motivated by the fact that it coincides with the *normative* task of (exhaustively) spelling out how we should conduct epistemic inquiry.

More specifically, we are provided with a rationale for the particular methodological approach of *Intuition-Driven Romanticism*, characterized by Weinberg as coming out of the idea that "the job of epistemologists is to get [the concepts and norms] out and set them out clearly" and that "the best way to do so is to pump our spontaneous judgments about applying or withholding terms of epistemic praise or blame to various hypothetical cases."[15] Furthermore, if Intuition-Driven Romanticism is (and has been) as widespread an approach as Weinberg *et al.* claims, it is easy to see why conceptual analysis has remained a preferred method even when divorced from Platonic essences and wedded to meanings: It presents a direct elucidatory route to the concepts and norms that we *should* abide by.

---

[14] See Weinberg, Nichols, and Stich (2001).

[15] Weinberg (2006, p. 29).

A question remains, however: Why should we assume (SYNC), i.e., that our epistemic concepts, norms, and goals are in full sync? Or put differently: Why assume that what we *should* do and what we (for whatever reason) are *prone* to do coincide? Well, for one thing, our belief-forming tendencies and strategies have been designed over millions of years of natural selection and, furthermore, a lifetime of learning. Surely this must have served to eliminate quite a few sub-optimal features of human cognition. Or as noted by Quine in an oft-cited passage: "Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind."[16]

More specifically, consider the following (admittedly informal) *reductio*: Say that our epistemic concepts, norms and goals are *out of* sync with each other in the sense that, when employing our epistemic concepts and adhering to our epistemic norms, we tend to *not* reach our epistemic goals. Under the plausible assumption that one central epistemic goal is true belief, it follows from our concepts, norms and goals being out of sync with each other that we tend to *not* have true beliefs when employing our epistemic concepts and adhering to our epistemic norms. Furthermore, since true (or at least approximately true) belief is an integral part of attaining most practical goals, including those involved in survival, it, furthermore, seems to follow that we tend to not survive—which is demonstrably false. Hence, we reject the initial assumption.

However, this *reductio*, at most, lends support to the idea that

> (SYNC*) our epistemic concepts and norms are sufficiently in sync with our epistemic goals to guarantee that, by adhering to the norms in which our epistemic concepts figure, we tend to have sufficient success in reaching our epistemic goals to guarantee the attainment of most practical goals.

[16] Quine (1969, p. 126).

More specifically, the above *reductio* does not show that our epistemic architectures are in *full* sync but only that they are not *radically* out of sync (i.e., that we are not *inveterately* wrong in our inductions, as Quine would say), which is completely compatible with the idea that

> (ALT) our epistemic concepts and norms are not necessarily in *full* sync with our epistemic goals, in the sense that there might be an *alternative* set of concepts and norms such that, if we were to employ those concepts and norms instead, we would reach our epistemic goals to a greater degree than we are currently doing.

Note that (ALT) should not be read as a mere modal claim about possible sets or concepts and norms (although such a reading is sufficient to drive a *conceptual* wedge between our *actual* concepts and norms and the concepts and norms that we *should* use). Instead, I will understand it in the stronger sense, as opening up for the potential *improvement* of our epistemic architectures, in not only entertaining the possibility that there might be an alternative set of norms and concepts that would enable us to reach our epistemic goals to a greater degree, but also that this set might just be *accessible* and *applicable*.

Let me illustrate this point by way of a hypothetical example: Say that an elucidation of our (or, at the very least, a prevalent) concept of justification yields the conclusion that to be justified is to reason in accordance with one's evidence in the specific sense of scrutinizing the evidential connections that hold between one's beliefs and their grounds and only assent to those propositions that survive such scrutiny. Assume, furthermore, that this concept is prevalent enough to warrant the claim that utilizing it tends to yield true belief to an extent that enables one to attain most practical goals to a sufficient degree. Now, consider the following empirical result: Our ability to scrutinize the evidential relations between our beliefs and their grounds are, in many situations, weakened by the dual fact that (*a*) we seldom have introspective access to the grounds for our beliefs, and, in the cases where we do, (*b*) we often misconstrue them in ways that

may be flattering to our self-images but that, nevertheless, makes for quite unreliable reasoning tendencies.[17] If so, however, there is an alternative way to conceive of justification, if only in the minimal sense of taking this empirical fact into account and, thereby, providing a more promising tool in the attainment of true belief.

This somewhat vague statement will be spelled out later, when we will delve into these and similar empirical results, as well as their implications for analysis, in chapter 6 below. For now, however, it suffices to note that the example serves to illustrate a possible scenario in which epistemic architectures that are not radically out of sync may still be improved by way of an alternative (and amended) concept, which is all we need to show in order to establish (ALT). And as such, (ALT) undermines Intuition-Driven Romanticism, in not only (*a*) calling for an empirical investigation into the merits of our current architectures, but also (*b*) directly discrediting an exclusive focus on the norms and concepts we currently employ, while ignoring the possible sets of norms and concepts that may serve us better in the pursuit of our epistemic goals.

This presents two problems for conceptual analysis, as traditionally construed. First, in raising the question why we should be exclusively concerned with our current concepts, it serves to question *Exhaustiveness*. Granted, the first step of analysis should be the uncovering of the cognitive architectures that are (for better or worse) used in the normative evaluations of epistemic inquirers, since it is exactly against the background of a clearer picture of these aspects that the epistemologist may improve on the architectural components that are not sufficiently in line with a specified set of epistemic goals. However, as will become more obvious in the next chapter, this neither implies that we are better off with exhaustive rather than approximate accounts, nor that epistemology should only be concerned with concepts, which brings us to the second point: improvements have to incorporate—indeed, be preceded by—an empirical evaluation of the

[17] See Wilson (2002).

epistemic merits of our architectures. Hence, Philip Kitcher, when discussing the failure of appeals to conceptual truths in epistemology:

> If an epistemological theory tells us that a particular policy of belief formation is justified or a particular type of inference is rational, and that these claims are analytic, that they unfold our concepts of justification and rationality, an appropriate challenge is always, "But why should we care about these concepts of justification and rationality?" The root issue will always be whether the methods recommended by the theory are well adapted for the attainment of our epistemic ends, and that cannot be settled by simply appealing to our current concepts.[18]

To sum up, two questions have been raised. First, given that there is no guarantee that our current architectures present the optimal pathway to our epistemic goals, what epistemological weight should be given to our current epistemic concepts? Second, if traditional conceptual analysis is not likely to do the trick, what is the role of more straightforwardly empirical investigation in epistemology? The following sections evaluate two methodological suggestions as to how to answer these questions. Both suggestions will eventually be deemed unsatisfactory. Chapter 3 constitutes my attempt to provide adequate answers.

2.5. Narrow Reflective Equilibrium

The previous section served to discredit the idea that *Exhaustiveness* is an important desideratum for analysis, although we will have to wait until the next chapter for a more conclusive argument against it. For now, we will instead look closer at an influential methodological suggestion that, unlike DCA and PCA, relinquishes *Exhaustiveness* in favor of the construction of philosophical theories that put our categoriza-

[18] Kitcher (1992, pp. 63-64).

tion intuitions and norms in a reflective equilibrium, i.e., in a state of stable and balanced co-existence.

The idea of a reflective equilibrium was introduced by Nelson Goodman—without using the label "reflective equilibrium," however—as a way to account for the justification of inductive norms.[19] More specifically, Goodman argued that inductive norms are justified in so far as they can be brought to cohere with a large set of particular judgments about acceptable inferences. However, this is not to say that our particular judgments hold any kind of privileged status. On the contrary, they can be rejected as misguided if shown to stand in conflict with general norms that we not only find acceptable but that may also account for a wide range of (other) particular cases. In other words, reflection and rejection goes both ways.

We will not be concerned with reflective equilibrium as a theory of justification for inductive norms. Rather, we will be concerned with reflective equilibrium as a desired end state of philosophical analysis. As such, the method of reflective equilibrium was introduced by John Rawls in his theory of justice.[20] According to Rawls, the appropriate way to formulate and choose among different conceptions of justice is in a situation in which our knowledge is constrained in very specific ways. For example, we are not to know the color of our skin, to what class or income bracket we belong, etc. Under such circumstances of constrained knowledge—in what Rawls called *the initial situation*—we would choose norms guaranteeing equal basic liberties to all, making sure that those that are worst off are as well off as possible.

However, Rawls also argued that we should not automatically accept the norms that, thereby, emerge. The norms must be in a reflective equilibrium with our considered judgments about justice. Hence, when searching for the most favored description of the initial situation we have to "work from both ends," as Rawls puts it:

---

[19] See Goodman (1983; originally published in 1955).

[20] See Rawls (1971).

> We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principles, I assume that eventually we shall find a description of the initial situation that both expresses reasonable judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium.[21]

More specifically, consider the following suggestion:

> *Analysis via Narrow Reflective Equilibrium (NRE)*
> For any analysandum '*F*,' reflect on (*a*) the logical and evidential interconnections among your intuitive categorization judgments regarding (being) *F*, as well as (*b*) any general norms that may be brought to bear on (being) *F*, and then construct a theory resolving any conflicts that are uncovered in the course of these reflections, so as to bring your beliefs and norms in a reflective equilibrium.

When such a reflective equilibrium has been reached, our norms and judgments coincide, and any "irregularities or distortions" in either have been ironed out in the name of coherence. It is important to stress that, in the pursuit of such coherence, our categorization intuitions, and the concepts that they can be taken to reveal, are no more

[21] Rawls (1971, p. 20).

holy than the norms we happen to adhere by. Consequently, just like a general norm may be reconsidered in light of a conflict with particular judgments to which we are strongly attached, a conflicting categorization judgment need not in all cases constitute a flaw in the theory, but in some cases rather be rejected, given that the norm responsible for the conflict can reasonably be deemed more central, explanatory, and hence, more important than that particular judgment. Hence, the rejection of *Exhaustiveness*.

However, NRE fails to take into account the particular reason discussed in the previous section as to why *Exhaustiveness* should be rejected, namely that there is no guarantee that our current norms and concepts are sufficiently in sync with our goals. NRE is a method by which we may clarify our stance and, in the process of doing so, bring our beliefs and norms into greater coherence. As such, it may amend conflicts between the particular and general judgments to which we are prone to assent. But the goal of analysis is not merely to prune our norms and judgments so as to fit together—it is also to make sure that our conceptual tools serve us well in the attainment of our goals. As construed above, the method of reflective equilibrium is not sensitive to the kind of mismatches that we discussed in the previous section and that motivated the rejection of *Exhaustiveness*. Merely bringing our general norms and categorization intuitions in a reflective equilibrium carries no promise to the effect that our norms or concepts are, thereby, altered in a way so as to be more in line with our goals.

2.6. Wide Reflective Equilibrium

Rawls shows some signs of being aware of this problem—or at least a close cousin of it. Later on in his book, he writes that there are "several interpretations of reflective equilibrium" and that

> […] the notion varies depending upon whether one is to be presented with only those descriptions which more or less match one's existing judgments except for minor discrepan-

> cies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them. In the first case we would be describing a person's sense of justice more or less as it is although allowing for the smoothing out of certain irregularities; in the second case a person's sense of justice may or may not undergo a radical shift. Clearly it is the second kind of reflective equilibrium that one is concerned with in moral philosophy.[22]

Similarly, Michael DePaul notes that the process of seeking reflective equilibrium, properly construed, cannot be a question of mere coherence:

> Even if the philosopher manages to bring her considered judgments and moral theory into a state of balance or [narrow] equilibrium via such a process of mutual adjustment, her work will not be finished. The philosopher must seek an even wider equilibrium. She must also consider the connections between her moral beliefs and principles and the other sorts of beliefs, principles and theories she accepts or rejects.[23]

This is also the move made by Norman Daniels when fleshing out Rawls claims in terms of so-called *wide* reflective equilibrium.[24] "The method of wide reflective equilibrium" Daniels explains, "is an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely (a) a set of considered moral judg-

---

[22] Rawls (1971, p. 49).

[23] DePaul (1998, p. 295).

[24] As noted above, the idea of wide reflective equilibrium seems to be present implicitly already in Rawls (1971). See also Rawls (1974/5, p. 8), where the idea is explicit.

ments, (b) a set of moral norms, and (c) a set of relevant background theories."[25] He continues:

> We begin by collecting the person's initial moral judgments and filter them to include only those of which he is relatively confident and which have been made under conditions conducive to avoiding errors of judgment. […] We then propose alternative sets of moral principles that have varying degrees of "fit" with the moral judgments. We do *not* simply settle for the best fit of principles with judgments, however, which would only give us *narrow* equilibrium. Instead, we advance philosophical arguments intended to bring out the relative strengths and weaknesses of the alternative sets of principles (or competing moral conceptions). These arguments can be constructed as inferences from some set of relevant background theories (I use the term loosely). Assume that some particular set of arguments wins and that the moral agent is persuaded that some set of principles is more acceptable than others […] We can imagine the agent working back and forth, making adjustments to his considered judgments, his moral principles, and his background theories. In this way he arrives at an equilibrium point that consists of the ordered triple (a), (b), (c).[26]

Naturally, the extent to which incorporating such arguments and theories into the pursuit of equilibrium may handle the kind of mismatch called attention to above depends on exactly how we understand the way in which they may or may not be responsive to the ends we are striving for. More specifically, putting a certain theory or argument on the scale that we are attempting to put in an equilibrium will only promote the construction of an analysans that serves us well if the theory or argument is sensitive to whether or not construing the analysandum in one way rather than another will promote the attain-

[25] Daniels (1979, p. 258).

[26] Daniels (1979, pp. 258-259).

ment of the goals in question. For this reason, we need to incorporate explicitly *empirical* theories about the world, since only such theories may serve as a benchmark for whether or not our current norms and concepts serve us well in the attainment of our goals.

In other words, consider the following suggestion:

> *Analysis via Wide Reflective Equilibrium (WRE)*
> For any analysandum '*F*,' reflect on (*a*) the logical and evidential interconnections among your intuitive categorization judgments regarding (being) *F*, (*b*) any general norms that may be brought to bear on (being) *F*, as well as (*c*) our current best theories either about or in any other way relevant to the phenomena at hand, and then construct a theory resolving any conflicts that are uncovered in the course of these reflection so as to bring your beliefs, your norms, and our current best theories about the world in a reflective equilibrium.

By incorporating not only categorization judgments and norms, but also our best current theories about, or relevant to, the phenomena in question, WRE leaves room for the evaluation of the extent to which our current concepts and norms do or do not provide us with the best tools to attain our goals. As noted by Timothy Williamson, however, the problem with a characterization like WRE is that it is sadly inadequate as even a summary description of the philosophical method of research.

> The question is not whether philosophers engage in the mutual adjustment of general theory and judgment about specific cases—they manifestly do—but whether such descriptions of it are sufficiently informative for epistemological purposes.[27]

[27] Williamson (forthcoming, chapter 7, pp. 35-36; page references refer to manuscript).

More specifically, as it stands, WRE is little more than a call for honest, intellectual inquiry, grounded in our best empirical theories about the world. As such, it, clearly, has a place in a sound, philosophical methodology. The only problem is that the main challenge does *not* lie in noting that philosophers, in constructing their theories, need to take into account not only our categorization intuitions and norms but also our best theories about the world, but in providing a story about exactly what role our categorization intuitions and norms should play in philosophical methodology, and how these intuitions and norms should be weighed against empirical evidence and our current best theories about the world, so as to yield an analysans that provides an improved tool in the attainment of our goals.

The next chapter considers one suggestion as to how this challenge should be met, in the form of Hilary Kornblith's recent arguments to the effect that our concepts are only relevant to philosophical inquiry in so far as they provide a set of paradigmatic examples, paving the way for empirical inquiry into their referents, where the only factor relevant to a concept's worth is the extent to which it provides an accurate story about its referent. My own suggestion is introduced in chapter 3 and fleshed out in chapter 4, and agrees with Kornblith in that empirical inquiry should play a substantial role in philosophical inquiry, but criticizes his particular theory for failing to leave room for certain types of conceptual improvement that can reasonably be assumed to be crucial in epistemology.

2.7. Conclusion

In the present chapter, a rectification of conceptual analytic methodology, working with prototypes rather than neat conditions, served to put the spotlight on the fact that philosophical analysis, as it has traditionally been conducted, provides an inferior methodology in the understanding of concepts, in light of the substantially more rigorous methods of empirical psychology. However, it was also shown that this issue becomes relevant only if we assume that giving exhaustive accounts of our concepts really is an important desideratum for analy-

sis. I ventured to suggest that it is not and that an investigation into our current epistemic repertoire is primarily interesting to the extent that it informs us of a starting point for potential improvement—a task that does not require exhaustive analyses.

So what does it require? A second rectification was considered, coming from the idea that proper philosophical theorizing consists in bringing our categorization judgments in a reflective equilibrium with our general norms and theories. However, under its most plausible construal, the suggestion, unfortunately, only amounted to the claim that philosophical theorizing should proceed along the lines of honest, intellectual inquiry, grounded in our best current theories about the world, while it answered none of the substantial questions regarding the proper place of categorization intuitions, concepts, norms and empirical evidence in epistemological theorizing. It will be the burden of the next two chapters to answer these questions, and it will be shown that the challenges posed by them call for something quite different from what has traditionally been thought of as conceptual analysis.

Chapter 3.
Epistemology and Empirical Investigation

Recently, Hilary Kornblith has argued that epistemological investigation is substantially *empirical*.[1] As we saw in the previous chapter, it is not necessary for an adherent of conceptual analysis to deny this. However, Kornblith's claim is not restricted to a question of epistemological *method* but also concerns the proper *target* of epistemological investigation, an important component in Kornblith's case being that knowledge—one of the main targets of epistemological investigation—is a natural kind and, hence, open to straightforward, empirical investigation, quite independently of any traditional semantic-philosophical inquiry into the concept KNOWLEDGE. In this sense, Kornblith puts forward an *external* objection to conceptual analysis, i.e., an objection questioning that we should analyze concepts in the first place.

In the present chapter, I will argue two things. First, I will show that Kornblith's claim about epistemology being a substantially empirical investigation is, in fact, not contingent upon the further and, admittedly, controversial assumption that all objects of epistemologi-

[1] See Kornblith (2007, 2006, and 2002).

cal investigation are natural kinds. Second, I will argue that, contrary to what Kornblith seems to assume, this methodological contention does *not* imply that there is no need for attending to our epistemic concepts in epistemology. Understanding the make-up of our concepts and, in particular, the purposes they fill, is necessary for a proper acknowledgement of epistemology's role in conceptual improvement.

This constitutes a rebuttal of his external objection in so far as it establishes that the analysis of epistemic concepts has an important part to play in epistemology. However, as we shall see in chapter 4, it does *not* constitute a defense of the entrenched picture of conceptual analysis, since conceptual improvement, as it turns out, calls for a kind of analysis very different from conceptual analysis as traditionally construed.

3.1. Whose Concepts?

Let us tackle the issue from this angle: An extremely valid question that is posed far too rarely within epistemology (if not analytical philosophy at large) is: When doing conceptual analysis, whose concepts are being studied? If philosophers are genuinely interested in studying any concepts but their own (and, perhaps, some of their colleagues'), it is somewhat surprising that the empirical methods of psychology have not yet found their way into the philosophy departments, if only in order to determine whether the studied concepts are shared by the folk, who they all too often are ascribed to.[2] In this respect, Frank Jackson's defense of conceptual analysis is admirably candid:

> I am sometimes asked—in a tone that suggests that the question is a major objection—why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious opinion polls on people's re-

[2] For a commendable exception, see the *empirical semantics* of Arne Naess (1938, 1953).

> sponses to various cases? My answer is that I do—when it is necessary.[3]

Unfortunately, however, Jackson's nod to empirical methods turns out to be somewhat half-hearted, when he, already in the next sentence, gives an example of the kind of polling he has in mind:

> Everyone who presents the Gettier cases to a class of students is doing their own bit of fieldwork, and we all know the answer they get in the vast majority of cases. But it is also true that often we know that our own case is typical and so can generalize from it to others. It was surely not a surprise to Gettier that so many people agreed about his cases.[4]

Interestingly enough, recent empirical evidence on the responses to Gettier cases tells a different story. According to a study by Weinberg, Nichols, and Stich, there is some reason to believe that there might just be a quite substantial variance when it comes to the categorization intuitions of different people when considering Gettier cases—a variance that could indicate a set of concepts more heterogeneous than epistemologists talking about "our" concepts have assumed.[5] Granted, Weinberg *et al.*'s study is far from conclusive. However, within the last years, Richard Nisbett—a prominent social psychologist that we will get more acquainted with below—has amassed a rigorous body of data indicating cross-cultural differences in thought patterns, and presented in his book *The Geography of Thought*.[6] It is in

---

[3] Jackson (1998, pp. 36-7).

[4] Jackson (1998, p. 37).

[5] See Weinberg, Nichols, and Stich (2001).

[6] See Nisbett (2006). It should be noted that none of the studies conducted by Nisbett and his colleagues indicate that these differences are *innate*. To the contrary, the differences in question are highly dependent on cultural factors such as social surrounding and upbringing and may, in some cases, even be so

the context of this research program and this body of data that the methodological implications of Weinberg *et al.*'s study should be evaluated. Understood thus, its results (however tentative) give us reason to ask that the "we" whose concepts are being illuminated by the categorization intuitions in question is specified, and that the supposition that there is a considerable overlap in the intuitive judgments of different people, at the very least, is argued for rather than simply taken for granted.

However, even if there, in fact, were a great uniformity among the categorization intuitions of the folk, and their intuitive judgments could be taken to reveal a set of philosophically uncontaminated epistemic folk concepts, it is not altogether obvious that such a set provides the best material for epistemological analysis. Hence, Kornblith:

> We do not go out of our way, in the sciences, to have observations made by individuals so ignorant of relevant theory that their corpus of beliefs contains no theories at all which might threaten to affect their observations. By the same token, one might think that, in philosophical theorizing, consulting the intuitions of the folk, who have given no serious thought to the phenomena of knowledge, justification, the good, the right, or whatever subject happens to be at issue, not only shields the resulting intuitions from the potential bad effects of a mistaken theory, but it also assures that the positive effects of accurate background theory cannot play a role. Those who have devoted a lifetime to thinking about knowledge and justification, for example, are certainly capable of making mistakes, and their theory-mediated judgments about these matters are certainly not infallible. But this hardly suggests that we should, instead, prefer the intuitions, uninformed by any real understanding, of the ignorant. The suggestion that we should at-

---

sensitive to context that a subject can be primed to switch back and forth between different ways of reasoning.

> tempt to capture pre-theoretical intuition, however, seems to privilege the intuitions of the ignorant and the naive over those of responsible and well-informed investigators. I cannot see why this would be a better idea in philosophy than it is in science.[7]

In short, ignorance is not an asset in the laboratory, nor should it be considered one in philosophical inquiry.[8] If such an analogy with natural science is to be upheld, however, one might justifiably wonder: Why should philosophy be concerned with concepts at all, rather than the real facts of the matter? As Kornblith writes:

> The uninformed observer and the sophisticated scientist are each trying to capture an independently existing phenomenon, and accurate background theory aids in that task. Experts are better observers than the uninitiated. If the situation of philosophical theory construction is analogous, however, as I believe it is, then we should see philosophers as attempting to characterize, not their concepts, let alone the concepts of the folk, but certain extra-mental phenomena, such as knowledge, justification, the good, the right, and so on. The intuitions of philosophers are better in getting at these phenomena than the intuitions of the folk because philosophers have thought long and hard about the phenomena, and their concepts, if all is working as it should, come closer to accurately characterizing the phenomena under study than those of the naive. So on this

---

[7] Kornblith (2007, p. 34).

[8] Of course, this is not to deny that *controlled* ignorance might sometimes be a scientific virtue, as in cases of double and triple blind studies. Clearly, there is a big difference between promoting general ignorance and promoting *controlled* ignorance, informed by knowledge about the best ways to avoid bias. And while the latter might, indeed, be a virtue in many cases, the idea here is that the former should never be deemed an asset—neither in science nor in philosophy.

> view the target of philosophical analysis is not anyone's concept at all. Instead, it is the category which the concept is a concept of.[9]

This line of reasoning is fleshed out in Kornblith's *Knowledge and Its Place in Nature*, where he makes an intriguing case for the reconceptualization of epistemological analysis from a largely non-empirical investigation (be it an *a priori* one or not) to a substantially empirical investigation, arguing that knowledge—one of the main targets of epistemological investigation—is a natural kind, open to straightforward empirical scrutiny.[10] Assuming that knowledge is not unique in this respect, which is an assumption that Kornblith, indeed, seems to make, consider the following reconstruction of his reasoning:

The Argument

(A) All objects of epistemological investigation are natural kinds.

(B) If (A), epistemological investigation is substantially empirical.

(C) Hence, epistemological investigation is substantially empirical (A, B, *MP*).

(D) If (C), a thorough understanding of our epistemic concepts, over and above the phenomena that they pick out, is irrelevant to epistemological investigation.

(E) Hence, a thorough understanding of our epistemic concepts, over and above the phenomena that they pick out, is irrelevant to epistemological investigation (C, D, *MP*).

As far as I know, Kornblith has never explicitly stated this argument. Still, I take it to provide one of the most reasonable rationales for Kornblith's more general claims about the implications of his results

---

[9] Kornblith (2007, p. 35).

[10] See Kornblith (2002).

concerning knowledge to epistemological analysis at large.[11] The plausibility of this interpretive claim should become more obvious as we proceed.

That being said, I will, in the following, scrutinize, qualify, and criticize *the Argument* in two steps. More specifically, §§3.2 through 3.4 will serve to contest (A) but defend (C) by showing that the latter premise is plausible even given that all objects of epistemological investigations are artifactual (or "socially constructed") rather than natural kinds. Then, in §§3.5 through 3.7, I will show that (E), nevertheless, does not follow from (C), since (D) is false and the claim that epistemological investigation is substantially empirical, hence, does *not* imply that an understanding of our epistemic concepts is irrelevant to epistemology.

3.2. On Natural Kinds and the Implausibility of Premise (A)

It should be beyond doubt that *the Argument* is valid. Indeed, it consists in two *modus ponens* arguments, where the conclusion of the first, i.e., (C), makes up the first premise of the second. However, I would like to contest its soundness. For one thing, it hinges on (A), i.e., the controversial assumption that all objects of epistemological investigation are natural kinds. As already mentioned, Kornblith has, indeed, argued that knowledge, as it is being studied by cognitive ethologists (cognitive ethology being the study of animal cognition), is a natural kind. However, the crucial question here is whether this claim may be generalized to other objects of epistemological study, so as to render (A) plausible.

To answer this question, we need to say something about what constitutes a natural kind. In the words of Kornblith, the underlying ontological assumption involved in postulating natural kinds is that "the world consists not merely of individuals but of kinds of individuals as well" and that "this division of the world into kinds is not of

[11] See, e.g., Kornblith (2006).

our own invention."[12] Hence, characterizing the world in terms of some kinds rather than others may not only be more or less *convenient*—i.e., serve our purposes to a greater or lesser degree—but also more or less *accurate*. By way of a positive example, we have good reason to believe that H_2O is a natural kind, independent of human thought and present long before chemistry had reached a level of sophistication enough to unveil and describe it. By way of a negative example, we have little reason to believe that the Aristotelian distinction between sublunary objects (objects inside the orbit of the moon) and superlunary objects (objects outside the orbit of the moon) corresponds to anything like a natural division in nature.

It should be noted already at the outset that this notion of a natural kind is different from the one popular during the hey-days of logical empiricism. In the early 20th century, many philosophers of science subscribed to the idea that physics gave a privileged description of the world in the specific sense that it was just a matter of time before the foundations provided by physics could unify the sciences. Paul Griffiths sums up the relevant changes in the notion of a natural kind as follows:

> The "unity of science" has dwindled to a minimal notion of supervenience—the world studied by economics or population biology does not change independently of the world studied by molecular biology or by microphysics. In this new philosophy of science the exception-ridden generalizations of many life and social sciences are seen as the only way to uncover some of the regularity inherent in natural processes. […] To reduce these sciences to their physical substrate is to eschew some epistemic access to that regularity. It is to know less about reality. [In the words of Richard Boyd,] [t]his new philosophy of science has given rise to an "enthusiasm for natural kinds" in many special sciences. […] Natural kinds are no longer conceived as the subjects of the fundamental laws of nature. They

[12] Kornblith (1993, p. 14).

> are simply nonarbitrary ways of grouping natural phenomena.[13]

In a sense, this "new" way of conceiving of natural kinds is more in keeping with an idea that stems back far further than the beginning of the 20th century. An arguable precursor (and oft-cited metaphor) can be found in Plato's method of "Collection and Division," on which we should take care to "see together things that are scattered about everywhere and to collect them into one kind (*mia idea*)" and then "cut the unity up again according to its species *along its natural joints*, and to try not to splinter any part, as a bad butcher might do."[14] A further example can be found in *An Essay Concerning Human Understanding*, where Locke distinguishes between nominal and real essences.[15] Nominal essences are the abstract ideas we associate with (general) names and these ideas may or may not correspond to real essences, i.e., the insensible structures responsible for the properties that we encounter in perception. Interestingly enough, Locke arguably considers the question of whether our nominal essences correspond thus to be impossible to answer, since the real essences are simply beyond our ken.[16] He thereby takes a moderately skeptical position about real essences, or what we today would call natural kinds. Already this position is quite daring, however, in light of the even more radically nominalist standpoint that there are *no* natural kinds whatsoever—that categorizations never reflect structures inherent in nature but merely more or less entrenched (yet arbitrary) ways of slicing up the world.[17]

The motivation for such a nominalist position is most plausibly that, given the general virtue of ontological parsimony, the burden of proof is on the natural kind defender to provide a convincing answer to the following question: What explanatory work would postu-

---

[13] Griffiths (1997, p. 213).

[14] The *Phaedrus* in Plato (1953, p. 265d-e; my emphasis).

[15] See Locke (1996, book III, chapter vi, 2).

[16] See Locke (1996, book III, chapter vi, 9).

[17] See, e.g., Goodman (1983).

lating natural kinds do that cannot be done just as well by assuming that kinds are nothing but creations of the mind?[18] Kornblith suggests an answer: Only if we assume that there are natural kinds can we explain the success of mature science. More specifically:

> If the scientific categories of mature sciences did not correspond, at least approximately, to real kinds in nature, but instead merely grouped objects together on the basis of salient observable properties that somehow answer to our interests, it would be utterly miraculous that inductions using these scientific categories tend to issue in accurate predictions. Inductive inferences can only work, short of divine intervention, if there is something in nature binding together the properties which we use to identify kinds.[19]

In other words, the best explanation of scientific success is that many (if not all) of the nominal kinds utilized in scientific prediction actually correspond to natural kinds.[20] As Kornblith puts it: "When a successful scientific theory quantifies over some sort of object, that is the most powerful evidence we may have that those objects genuinely exist."[21]

---

[18] In fact, some of Locke's own writings seem to suggest this very line of reasoning. See his (1996, book III, chapter vi, 4).

[19] Kornblith (1993, pp. 41-42).

[20] This implication from success to realism, together with the more general idea that explanatory success provides evidence for the existence of the phenomena postulated in the explanation, has been contested in the literature on scientific realism, perhaps most famously by Arthur Fine (1986) and Bas van Fraassen (1980). However, since the aim of this paper is not to defend Kornblith's argument for the existence of natural kinds, I will not delve into this debate here. Still, I refer the reader to Ahlström (2006), where I express my doubts about the viability of Fine and van Fraassen's position, especially in light of an inability on their part to account for scientific *failure*.

[21] Kornblith (1993, p. 55).

Why is this so? The reason, Kornblith suggests, is the following:

> *Identity Conditions for Natural Kinds*
> The identity conditions for natural kinds are given by clusters of homeostatically related properties, i.e., properties that "when realized together in the same substance, work to maintain and reinforce each other, even in the face of changes in the environment."[22]

By virtue of the fact that natural kinds comprise such homeostatically related properties, it is possible to reliably infer the presence of some properties from the presence of others. Take water, for example. Water consists of molecules of two hydrogen atoms connected to an oxygen atom. Moreover, this particular chemical constitution is responsible for a wide range of other properties, such as being transparent, potable, a good solvent, and a compound that boils at 212 degrees Fahrenheit (under standard pressure). As a consequence, these and other characteristic properties of water form a homeostatic cluster, which is exactly why we may reliably infer a rich variety of properties and facts from knowing that we are interacting with water, rather than with some other substance or motley collection of properties.

Similarly, a bird is any member of the evolutionary branch—or *clade*, as biologists say—*Aves*. By sharing a common ancestry, members of *Aves* also tend to share a series of properties from the molecular to the behavioral level, which is exactly why we may reliably infer a multitude of properties from knowing that we are interacting with a bird, regarding everything from what toxins it will metabolize to what learning algorithms it will employ. Put differently, natural kinds are *projectible* in the sense that observations about them may be projected onto new instances. More importantly, according to Kornblith, the categories of water and bird are not unique in this respect. In actual-

[22] Kornblith (1993, p. 33).

ity, this generalizes to all natural kinds, which explains why latching on to natural kinds in prediction tends to yield predictive success.[23]

Having thus shed some light on what it is to be a natural kind, is it plausible to assume that all objects of epistemological investigation are natural kinds? Take epistemic justification, for example—a phenomenon that has been the subject of extensive scrutiny within contemporary epistemology. What are the prospects for extending Kornblith's case for knowledge to justification? Unfortunately, unlike KNOWLEDGE, JUSTIFICATION is not an entrenched concept in cognitive ethology. Hence, it is questionable whether Kornblith's case for knowledge can be extended to justification in any straightforward way. In fact, it is hard to see exactly how JUSTIFICATION, together with such related concepts as EVIDENCE, UNDERSTANDING, and RATIONALITY, at all *could* correspond to natural rather than artifactual (or "socially constructed") kinds, the latter of which do not comprise homeostatically structured conglomerates of properties independent of human understanding, but grids whose structure reflects nothing but human intentions (in a sense that will be elaborated on below). Still, as has been noted by Alvin Goldman and Joel Pust, the lack of natural kind status hardly places the topic of justification (or that of evidence, understanding, and rationality) outside the scope of epistemological analysis.[24]

So, on pain of radically restricting the scope of epistemological analysis (an option that I will not consider), the defender of *The Argument* has to face up to the following problem:

> *Problem 1*
> Unless (A) holds, there is little reason to believe that the applicability of epistemological analysis stretches beyond the analysis of one particular object of epistemological investigation,

---

[23] See also LaPorte (2004), who defines natural kinds as kinds with a high explanatory value in science—a definition that seems to yield roughly the same extension as Kornblith's.

[24] See Goldman and Pust (1998, pp. 186-187).

> namely knowledge. And even this particular application is contingent upon the admittedly controversial claim that knowledge, in fact, is a natural kind.

The following two sections discuss two solutions to *Problem 1*, both of which amount to the claim that there is a case to be made for extending the conception of epistemological analysis as substantially empirical to the analysis of artifactual kinds.

3.3. A First Attempt to Save (C): Content Externalism

The first solution starts out with the observation that it might be plausibly argued—and, indeed, has been argued by Putnam and, more recently, by Kornblith[25]—that content externalism provides the correct semantic not only for natural kind terms but also for artifactual kind terms. Rather than directly contesting this line of argument, the second solution (which is the one I will favor) concludes that, as it turns out, the plausibility of extending the claim about empirical analysis to artifactual kinds is largely independent of which semantic theory one accepts for the latter. Before evaluating any of these solutions, however, we need to say something about what constitutes artifactual kinds and, in particular, what distinguishes them from natural kinds.

To a first approximation, we may characterize artifactual kinds negatively as *not* comprising homeostatically clustered properties. However, even disregarding the fact that this characterization is hardly informative, it does not even uniquely pick out artifactual kinds, unless natural and artifactual kinds exhaust the realm of kinds (which they do not). So, by way of a positive characterization, we may say that artifactual kinds are somehow dependent on human intentions. However, this formulation is not only vague but also potentially misleading if not further qualified. Take polyethylene or amphetamine, for example. Since they are synthetic substances, it is plausible

[25] See Putnam (1975b) and Kornblith (forthcoming a).

to assume that neither of them would be around if it were not for certain human intentions, pertaining to the need for a light, flexible, yet tough material or a substance to fight fatigue and increase alertness among servicemen.

Still, this only serves to show that the existence of some (instances of) synthetic substances is *casually* dependent on certain human intentions. It does not show, however, that the kinds to which those substances correspond are *ontologically* dependent on human intentions. That is, it does not show that the *identity conditions* for polyethylene or amphetamine—i.e., the conditions specifying what *makes* something an instance of polyethylene or amphetamine—are in any interesting sense intertwined with human intentions. Indeed, an acknowledgement of this very fact is implicit in what we take to be the best explanation of why polyethylenes and amphetamines fit into reliable inductive generalizations better than any random motley of properties. This explanation assumes that polyethylenes and amphetamines are endowed with an underlying chemical composition (i.e., C_2H_4 and $C_9H_{13}N$, respectively), and that this, furthermore, accounts for the fact that some inductions involving the respective substances are successful (e.g., from "this is amphetamine" to "this will increase stamina but decrease appetite if ingested") while others are not (e.g., "this is made of polyethylene" to "this is blue"). Hence, they may plausibly be considered natural kinds.

Not so for, say, pens—a clear example of an artifactual kind. There is no need to assume that all pens share an underlying nature to explain why certain inductions involving pens are successful (e.g., from "this is a pen" to "this can be used to write with") while others are not (e.g., from "this is a pen" to "this is warm"). The reason is that instances of artifactual kinds owe their kind membership exclusively to the fact that they fulfill certain purposes. More specifically, I suggest the following:

Identity Conditions for Artifactual Kinds
The identity conditions of artifactual kinds are given by sets of human intentions, pertaining to the fulfillment of certain purposes.[26]

Clearly, this is not to say that artifactual kinds *consist of* sets of human intentions and purposes, but that what determines whether or not something is an instance of a particular artifactual kind pertains to whether that something can fulfill certain purposes and, thereby, answer to a specific set of human intentions.[27] Thus, a pen is a pen (roughly) by virtue of fulfilling the purpose of drawing and writing and, thereby, answering to certain human intentions regarding creative outlet and communication, just like a key is a key (roughly) by virtue of serving the purpose of locking and unlocking doors, lockers, etc., and, thereby, answering to a set of human intentions regarding controlled access to certain spaces.[28]

Let us now consider Kornblith's claim that the semantic mechanisms of reference for artifactual terms are insensitive to these ontological differences between natural and artifactual kinds.[29] Take an SUV, for example—clearly, an example of an artifactual kind. Unlike the case of water and polyethylene, there is no reason to assume that SUVs share a hidden nature, since an explanation of why

---

[26] See Thomasson (2003) for a more thorough treatment of a suggestion along these lines.

[27] It might be argued that it, for some artifactual kinds, is not sufficient for kind membership that something merely *can* fulfill certain purposes and, thereby, answer to certain human intentions, but that it also has to have come about as the result of an *intention* to fulfill those purposes. See Thomasson (2003, p. 594) for a discussion.

[28] I am not suggesting that such sets of intentions can be summed up in anything like a clear-cut conjunction of properties. This picture—just like actual categorizations of artifacts—is fully compatible with conceptual fuzziness and in-between cases.

[29] See Kornblith (forthcoming a).

we categorize the world and successfully reason in terms of SUVs and non-SUV type vehicles does not need to go beyond factors pertaining to certain (potentially superficial) properties regarding form (e.g., relative size) and function (e.g., performance), answering to certain human intentions concerning traveling and transportation. In fact, I am, personally, not sure what makes something an SUV and, in particular, not what distinguishes it (if anything at all) from a jeep, van or any other fairly big motor vehicle with four wheels. Regardless of whether I, thereby, just happen to be exceptionally uninformed concerning motor vehicles, however, I take it that I, nevertheless, just succeeded in referring to SUVs. How can that be?

Perhaps it is due to the dual fact that (*a*) there are people in my linguistic community that *do* know what makes something an SUV and (*b*) my successful reference to SUVs is parasitic upon their knowledge and ability to discriminate SUVs from non-SUV type vehicles. But are these conditions *necessary* for successful reference? Is it, in particular, necessary that there is at least one member of my linguistic community that knows what, thereby, constitutes SUVs? Remember that what makes something an SUV pertains to a set of human intentions and purposes—not anything like an underlying nature, shared by all SUVs. So, is successful reference contingent upon there being at least one member of my linguistic community that knows what this set is, i.e., to what intentions SUVs need to answer and what purposes they need to fulfill? Consider the following two responses, corresponding to two variants of content externalism:

> *Social Externalism about Artifactual Kind Terms*
> Successful reference to artifactual kinds (only) requires that there is at least one member of the relevant linguistic community (i.e., an "expert") that can correctly delineate the set of human intentions and purposes that provides the identity conditions for the kind in question. Subsequent instances of suc-

cessful reference to this kind are then parasitic upon the discriminatory competence of this member.[30]

Causal Externalism about Artifactual Kind Terms
Successful reference to artifactual kinds (only) requires that a sample of the kind has been picked out in an initial act of baptism through an ostensive definition, fixing the set of human intentions and purposes that provides the identity conditions for the kind in question and establishing a socially sustained chain of reference upon which subsequent instances of successful reference to whatever bears a certain equivalence relation (perhaps spelled out in terms of certain potentially superficial properties regarding form and function) to the ostended sample are parasitic.[31]

Clearly, neither formulation is intended to constitute a full-fledged theory. If anything, they both give rise to further questions. Take *Causal Externalism*, for example. For one thing, it is somewhat puzzling how the mere causal relation involved in an act of baptism could fix a *unique* set of human intentions and, consequently, pick out a *single* kind. For example, take a wooden, box-like, and fairly heavy object of 35 by 35 inches that we may name *b*. Picking out *b* as a sample for an artifactual kind, what determines the relevant set of intentions, given that *b*, among many other things, can be used to sit on (i.e., as a chair), to sit by (i.e., as a table), to stop doors from closing (i.e., as a doorstop), or, to stop people from ascending (not to mention hurt them quite badly in the process), if the box is pushed down a set of stairs?

One plausible suggestion is that there is something about the mental state of the baptizer that determines the relevant set of intentions, namely that the mental state (at the very least) *instantiates* that very set of intentions. If so, however, it becomes harder to distinguish *Causal* from *Social Externalism*. For example, is successful reference

[30] Cf. Burge (1986).

[31] Cf. Kornblith (forthcoming a).

contingent upon (*a*) the act of baptism and the resulting chain of reference, or (*b*) the fact that there is a baptizer, carrying the heaviest burden in the division of linguistic labor due to her insight into the relevant set of intentions (granted introspective access, of course)? It is not so clear to me which one is the case here.[32]

However, the relevant question for our purposes is how to *analyze* artifactual kinds and, in particular, whether any of the above considerations render the claim that epistemology is substantially empirical implausible. They do not. The idea that epistemology is substantially empirical can be plausibly extended to artifactual kinds, regardless of whether *Social* or *Causal Externalism* holds and for the following reason: Both *Social* and *Causal Externalism* are fully compatible with successful reference despite substantial ignorance regarding many properties of the entities or phenomena picked out. Granted, *Social Externalism* implies that there is at least one member of the relevant linguistic community that has insight into the identity conditions of the kind in question. So, in the general case, the baptizer, clearly, knows *something* about the artifactual kind she is baptizing, such as that it can be used for certain purposes and, thereby, answer to certain human intentions. However, her knowledge of many of the properties that make up instances of the artifactual kind in question may be ever so limited—or better said: there is nothing in her role as a *baptizer* that hinders her knowledge from being limited thus. That is, even if we reject *Causal* in favor of *Social Externalism*, the *epistemically most privileged user* of an artifactual kind term, i.e., the baptizer herself, may have an ever so limited insight into the properties that do or may make up instances of the kind in question.

As the reader surely suspects, it is exactly in this potential gap between successful reference and insight into the properties of the referent picked out that our first solution to *Problem 1* gets its foothold, since such a gap makes possible scenarios in which (*a*) a majority of speakers either are largely *ignorant* of or have a highly *inaccurate*

---

[32] See Stanford and Kitcher (2000) for a discussion of this and similar problems.

conception of many of the properties that make up the instances that they are successfully referring to and (*b*) even the most epistemically privileged speaker (i.e., the baptizer) may have an ever so limited insight into the multitude of properties that may not in any straightforward way be inferred from the intended purpose of the kind in question. Given such a gap, I take it that it would reasonably follow that epistemology is substantially empirical, even given that most (if not all) objects of epistemological investigation are artifactual kinds.

3.4. A Second Attempt to Save (C): Conceptual Refinement

The problem with this solution, however, is that it is far from controversial that anything like *Social* or *Causal Externalism* provides the correct semantics for artifactual kind terms. While remaining essentially neutral on this particular issue, and in an attempt to develop a somewhat more dialectically robust rationale for (C), I will now argue that, even if a strong form of *internalism* turned out to provide the correct semantics for artifactual kind terms, this would in no way undermine the claim that epistemology is substantially empirical. The argument will also indicate that it was not externalism that did the job in the above solution after all.

So, consider the following:

> *Strong Internalism about Artifactual Kind Terms*
> Successful reference to artifactual kinds requires that the speaker can correctly delineate the set of human intentions and purposes that provides the identity conditions for the kind in question.

Strong Internalism makes up the other extreme of the semantic spectrum; it is not enough that there is an appropriate causal chain of reference, nor that *someone* in the linguistic community can delineate the set of intentions in question—the speaker must *herself* be able to make such a delineation for her to successfully refer. Perhaps this is a more plausible thesis about the semantics of artifactual kinds, or per-

haps it is not. What is important to note for our purposes is that even if *Strong Internalism* turned out to be true, this would only imply that every competent user of artifactual kind terms were in the same epistemic situation as the epistemically most privileged user in the *Social Externalism* scenario. That is, while being extremely informed as to the relevant set of human intentions and purposes, they may still, *qua* competent users, have an ever so limited insight into many of the properties that make up actual instances of the kind in question. More importantly, they may, in the epistemic case, be ever so uninformed concerning properties of epistemological *significance*—or so I will now argue.

First, consider the following definition:

> *Conceptual Accuracy*
> A concept is *accurate* to the extent that it provides a *correct* and *complete* description of its referent.

The idea here is two-fold: (*a*) There are aspects of concepts that do not serve to determine reference, and (*b*) these aspects may be represented in terms of descriptions. Let us look closer at (*a*) first. As was noted above, we may distinguish between factors that serve to *fix* reference and factors that serve to *determine* reference.[33] For example, while whatever conceptual component responsible for my tendency to think of horses as having four legs may serve to *fix* the reference of HORSE—i.e., to be a helpful tool in picking out actual horses in my environment—it does not *determine* reference, for the simple reason that some horses are amputees. As such, factors fixing reference, clearly, play an important cognitive role in our mental life, by significantly facilitating our interaction with the extra-mental world. *Contra* the descriptivist, however, they should not be confused with the factors determining reference. We will return to this point below.

Let us now turn to (*b*). If no conceptual aspect could be represented in terms of descriptions, it is hard to see how concepts at all

[33] See Kripke (1980, pp. 55, 57, and 96).

could be the objects of any kind of analysis in the first place. On any view of concepts—be it concepts as abilities, Forms, senses or mental representations—concepts serve to categorize the world, and the systems of categorization that arise from concept use may be represented in terms of descriptions. Hence, a concept that serves to put all and only blue objects that weigh more than two pounds in one category may be characterized in terms of the description "is blue and weighs more than two pounds," quite independently of one's favored ontology of concepts. This is *not* to say that whatever mental occasion that is causally responsible for the categorization takes the form of a description—that would have to be established through empirical research—which is exactly why I am not saying that concepts *are* descriptions but merely that certain aspects of a concept for present purposes may be *represented* in terms of descriptions.

Reluctance to talking about concepts and descriptions in the same sentence typically stems from an aversion to the aforementioned idea that referents are determined by concepts by virtue of the former satisfying descriptions inherent in the latter. However, as should be clear by now, this is certainly not the idea being defended here. If anything, the present notion of conceptual accuracy serves to state in clearer terms the very externalism that served to refute this descriptivist picture of the factors determining reference: In so far as any form of content externalism holds, having an accurate concept—i.e. a concept providing a correct and complete description of its referent—is *not* a prerequisite for successful reference. However, since we are, for the moment, assuming that *Strong Internalism* provides the correct semantics for artifactual kind terms, we will focus on the fact that not even *Strong Internalism* implies that having an accurate concept of an artifactual kind is a prerequisite for successfully referring to it. The reason may be brought out as follows:

Take any object x of epistemological investigation. If x is an artifactual kind, there is a set T of human intentions and purposes that determines the identify conditions for x. What will T contain? Given that x is an artifactual *epistemic* kind, it will most likely contain intentions and purposes pertaining to the attainment of certain epistemic

goals, say, true belief in significant matters. (We will return to the exact make-up of our epistemic goals in chapter 4 below.) Is it possible to say something more specific? Well, if *Strong Internalism* holds, we may note that, unless we want to radically restrict the extent to which people may successfully refer to *x*, we have to restrict the richness of information contained in *T* since, in the general case,

> the more information is contained in *T*, the more rare is successful reference to *x*.

Given that reference to *x* is widespread, however—which is, hopefully, the case for most epistemic kinds—we may, at the very least, say that

> being acquainted with the information contained in *T* cannot involve a complete knowledge of all properties that make up instances of *x*.

Hence, even given *Strong Internalism*, the semantically most informed person—i.e., the person that has the most complete grasp of what is contained in the set *T*, that determines the identify conditions—might still be in the dark as to many of the properties that make up instances of *x*. In other words, she may still have an inaccurate concept of the kind in question.

I take it that few would deny this claim, if understood in the weak sense of there always being further facts that could be found out that do not flow from what we know just by virtue of being able to successfully refer. For example, just by virtue of successfully referring to pens and keys, I may (at least on the *Strong Internalist*'s story) know a whole host of things about pens and keys and, in particular, things that I may easily infer from being acquainted with the relevant sets of intentions and purposes. At the same time, there may very well be a lot of things that I do *not* know about pens and keys, such as its exact mechanical make-up, its molecular constitution, etc. As pointed out by Paul Griffiths, this is to be expected considering that the factors

determining the identity conditions are substantially less rich in the case of artifactual kinds than in the case of natural kinds:

> The traditional natural kinds are among the richest. The kindhood of a physical element determines almost all its salient properties. […] In contrast, knowing what sort of thing an artifact is, knowing that it is a bracelet for example, may fix very few of its features. There are just too many ways to skin a cat, or in this case too many ways to decoratively encircle the wrist.[34]

Of course, it does not follow that being familiar with the multitude of ways that pens or keys may be crafted or wrists may be decoratively encircled is necessarily very *significant* to me, given that my goals, as far as pens, keys, and bracelets go, are restricted to successful, everyday interactions. What I want to claim, however, is that the same does *not* hold in the epistemological case. In particular, I want to claim that what is not contained in *T*—i.e., what is still to be found out when we have enough knowledge for successful reference to occur—is of *epistemological significance*.

My argument for this claim runs as follows. First, remember what was said in the previous chapter about the improvement of epistemic architectures. The underlying rationale for taking such improvements to constitute an important part of the epistemologist's job description is the idea that epistemology should *guide* epistemic inquiry. This idea will be further elaborated on in the second part of this study, but we may note already at this point that one important component to this element of guidance is that epistemology should see to it that our epistemic vocabulary is as *apt* as possible, where a vocabulary is apt in so far as it invokes apt concepts, and

[34] Griffiths (1997, p. 190).

Conceptual Aptness
a concept is *apt* to the extent that it serves its intended purpose well in use.

To say that concepts serve purposes is not meant to imply anything controversial. At a very basic level, the purpose of concepts is simply to enable us to think certain thoughts, have certain beliefs, etc., and, thereby, interact with the world in more or less successful ways. As noted by Armstrong (following Ramsey), beliefs are the maps by which we steer.[35] Clearly, some ways of steering are more successful than others. That is, some ways of steering will get us what we want to a greater extent than others. Furthermore, one important factor in successful intellectual navigation are the concepts used, since concepts provide *frameworks* for thinking and believing, in the sense of different ways of categorizing the world. This point may be illustrated in the epistemic domain by noting that we, as epistemic inquirers, are engaged in a certain project of epistemic evaluation and doxastic revision, (roughly) aimed at attaining and maintaining true belief in significant matters. An integral part of succeeding in this latter task is having an apt epistemic vocabulary, where an apt epistemic vocabulary is a vocabulary that can be used to categorize and, thereby, evaluate fellow inquirers and the world in a way that serves the purpose of attaining and maintaining true beliefs in significant matters. Clearly, some concepts will serve this purpose better than others. In particular, the following seems a reasonable claim:

> If epistemic vocabulary V_1 is more *refined* than vocabulary V_2—i.e., if V_1 incorporates *accurate* concepts to a larger extent than V_2 does—then V_1 is more *apt* than V_2, *ceteris paribus*.[36]

---

[35] See Armstrong (1973) and Ramsey (1931).

[36] I will attend to the issues about the identity conditions of concepts—i.e., whether and to what extent a refined concept can be said to remain "the same" over the course of refinement—more fully below when discussing conceptual reconstruction.

Furthermore, it is reasonable to assume that getting acquainted with the properties that make up actual instances of *x*—and, in particular, those properties that may not be readily inferred from being acquainted with what is contained in *T*—would enable us to refine our concept of *x*, in the sense of pruning it so as to provide a more correct and complete description of its referent. Finally, it seems fairly uncontroversial that the proper method for getting acquainted with those properties will have to be empirical.

However, some might be skeptical as to whether we thereby get to know something that adds to our *concepts*. More specifically, it might be argued that there is no doubt that you may come to know all kinds of interesting and maybe even significant things by way of an empirical investigation into referents. However, the things you, thereby, come to know do not (at least not in all instances) add to the *concept* in question—only those things that are necessarily true do.

As we shall see in the next chapter, this is false if we understand the relevant conceptual aspect in terms of so-called stereotypes.[37] For one thing, the predicates that make up stereotypes do not even need to be *true* of the referent (let alone *necessarily* true). To see this, we need to remember two things, namely that (*a*) an important reason to postulate concepts is to explain the ways in which we think and act, and that (*b*) mistaken theories and, hence, stereotypes for WATER and ATOM in no way hindered pre-1750 chemists and late 15th century physicists from picking out certain very real physical phenomena in nature that not until centuries later came to be accurately described. Given (*a*) and (*b*), it would be an explanatory disaster to only allow truths—let alone *necessary* truths—as conceptual components. The particular ways in which pre-1750 chemists tended to interact with water, and late 15th century physicists tended to interact with atoms, are best explained not in terms of any perfectly accurate theories of the phenomena in question, but by the actual stereotypes to

[37] See Putnam (1975a, 1975b).

which they corresponded, however inaccurate they may have turned out to be.

Even this point aside, however, we may note that there are independent reasons to assume that not even those stereotypical components that happen to be true of the referent need to be *necessarily* true. For example, take the fact that human parents tend to care for their offspring. Clearly, not all parents care for their offspring. Nevertheless, it seems that most parents, at the very least, have a *tendency* to care for their offspring. However, this is not a necessary property of parents (unlike, say, that all parents are older than their offspring). Given a slightly different evolutionary history, we might have had a substantially different relationship between parents and offspring, such that the former had no tendency whatsoever to care for the latter. Nevertheless, it would be wrong to say that it is not part of the stereotype of HUMAN PARENT that parents tend to care for their offspring. When we think of parents and their offspring, we think of the former as having a tendency to care for the latter. And there are good reasons for doing this; in the actual world—the world that we live in and interact with—parents tend to care for their offspring. So, to the extent that our concepts figure in the ways that we interact with and explain the world, and stereotypes provide viable ways to represent certain cognitive aspects of such thought and interaction, it serves us well to incorporate "…tends to care about their offspring" in our stereotype for HUMAN PARENT.

By way of another example, take the fact that many birds migrate south in the winter. This is not a necessary property of these birds (unlike, say, that all and only birds are members of the biological clade *Aves*). If there were not seasonal changes in the climate, there would be no reason for birds to migrate south in the winter. As it happens, however, there are seasonal changes and these very seasonal changes make certain birds migrate south during the summer. Hence, when we think of these birds, we tend to think of migration—at least if we are somewhat informed as to the habits of birds. And there are good reasons for thinking about birds thus; in the actual world—a world with seasonal changes and birds that are sensitive to them—

many birds migrate south during the winter. Hence, to the extent that our concepts figure as prominent tools in the ways that we interact with and explain the world, and stereotypes can be taken to represent important cognitive aspects of those tools, it serves us well to incorporate "…migrates south in the winter" in our stereotype for BIRD.

Parents, birds and other natural kinds aside, let us turn to justification, by way of epistemological illustration. Construed as an artifactual kind, the set of intentions and purposes that determines its identity conditions would most likely pertain to the flagging of appropriate sources of information, where the appropriateness is understood in terms of truth-conductivity. (Or so I will argue in chapter 7 below.) More than that, on *Strong Internalism*, every competent speaker would be perfectly familiar with the details of this set. But does this imply that they, thereby, know everything there is to know about justification? Does it, in particular, follow that there is nothing else to find out that is of *epistemological significance*?

That seems unreasonable. In particular, it would be of epistemological significance to find out, among other things, (*a*) what external phenomena actually satisfy the relevant requirements of truth-conductivity, (*b*) about the multiplicity of properties that make up these phenomena, and, perhaps more importantly, (*c*) how these properties fit into the causal fabric of the world and, hence, may not only be better understood but also be manipulated to the benefit of the epistemic inquirer—all of which seem to be things that cannot, in any straightforward way, be inferred from the relevant set of intentions and purposes, nor be discovered without recourse to an empirical investigation. Furthermore, given that we drop the implausible requirement that only necessary truths may figure as conceptual components, there is nothing that hinders the results of such investigation from being incorporated into an increasingly accurate and, hence, more useful concept of justification.

So, in the general case, and in so far as conceptual *refinement* may rightfully play a substantial role in epistemology, the claim that epistemological investigation is a substantially empirical investigation is largely independent not only of whether the objects of epistemo-

logical investigations are natural or artifactual kinds, but also of any content externalist or internalist considerations with respect to the latter. In other words, given that the above line of reasoning does, indeed, apply to most if not all objects of epistemological investigation—and I see no reason why it would not—we may conclude that the claim that epistemological analysis is substantially empirical does not hinge on (A). Indeed, the preceding discussion provides us with reason to take the following argument to be sound:

(A*) For every object x of epistemological investigation, x is either a natural or an artifactual kind.
(B*) If (A*), epistemological investigation is substantially empirical.
(C) Hence, epistemological investigation is substantially empirical (A*, B*, *MP*).

This concludes my solution to *Problem 1*. Next, we will be looking into why (D) does *not* hold—i.e., why it does not follow from (C) that an understanding of our epistemic concepts is largely irrelevant to epistemological investigation. In the process of doing so, we will not only provide further evidence to the effect that epistemology is substantially empirical, but also introduce a more radical means to attaining an apt vocabulary: conceptual reconstruction.

3.5. Factual Analysis and the AC Principle

Returning to *The Argument*, let us now turn to premise (D) and the conclusion (E), stating that a thorough understanding of our epistemic concepts, over and above the phenomena that they pick out, is irrelevant to epistemological investigation. Kornblith's commitment to this conclusion comes out most clearly in his critique of the idea that the (sole) job of epistemology is to analyze epistemic concepts by way of categorization intuitions about hypothetical cases—a view that he characterizes as follows:

> Appeals to intuition are designed to allow us to illuminate the contours of our concepts. By examining our intuitions about imaginary or hypothetical cases, we should be able to come to an understanding of our concepts of, for example, knowledge and justification. The goal of epistemology on this view, or, at a minimum, an essential first step in developing an epistemological theory, is an understanding of our concepts.[38]

Kornblith continues:

> My own view is that our concepts of knowledge and justification are of no epistemological interest. The proper objects of epistemological theorizing are knowledge and justification themselves, rather than our concepts of them.[39]

In light of the reasonable claim that *some* initial examination of our epistemic concepts might be necessary in order to fix the subject matter, Kornblith makes it clear that his main disagreement with the tradition of epistemology as conceptual analysis concerns the scope of such a semantic investigation. More specifically, he claims that the semantic investigation called for is "utterly trivial" and, thereby, in no way related to the two thousand year old project that, in a tradition stemming from Plato's *Theaetetus* and culminating in the Gettier-inspired literature, typically falls under the heading of the analysis of knowledge and justification.[40]

When trying to put this claim in more precise terms, it serves us well to make a distinction between two stages of epistemological investigation. The first one corresponds to the *identification* of an epistemological object *F*, i.e., of fixing the subject matter (if only tentatively) through picking out a selection of what we take to be paradigmatic instances of '*F*.' In doing this, our concepts of *F*, and the cate-

---

[38] Kornblith (2006, pp. 11-12).

[39] Kornblith (2006, p. 12).

[40] See Kornblith (2006, pp. 12-13).

gorization intuitions they give rise to clearly play a vital role. Moreover, depending on the extent to which '*F*' is ambiguous or imprecise, this process of identification may be more or less time-consuming. Regardless, the purpose of identification is to pave the way for the more substantial and straightforwardly empirical *aggregation* of the characteristics that are found in *F*s, in order to reach a satisfactory answer to the question that initiated inquiry in the first place: "What is (it to be) *F*?"[41]

Against the background of this distinction, I would like to characterize Kornblith's notion of analysis as follows:

> *Factual Analysis (FA)*
>
> *Identification*: For any '*F*,' Identify a set *Q*, containing a selection of what we take to be paradigmatic instances of '*F*.'
>
> *Aggregation*: Against the background of an empirical investigation into the elements found in *Q*, aggregate a set of characteristics that specify what actually constitutes (being) *F*.

Is FA an empirical analysis? In so far as aggregation goes, the answer would have to be yes. As stated, however, it remains to be established whether the same goes for identification. We saw in the previous chapter that the use of categorization intuitions are best understood as yielding *a posteriori* warranted beliefs by way of either an empirical (third-person approach) or a non-empirical method (first-person approach). Furthermore, and in light of the previous two chapters, it seems reasonable to assume that either of the following two claims would have to be true: (*a*) identification calls for an extensive investigation, in which case an empirical method will be desirable due to a

[41] This distinction between identification and aggregation is borrowed from Amartya Sen's (1981) excellent treatment of the issue of poverty.

superior methodological rigor,[42] or (*b*) identification does *not* call for an extensive investigation, in which case a non-empirical method would not compromise the claim that epistemological investigation—if understood along the lines of FA—is still a *substantially* empirical investigation.

We have to keep in mind, however, that Kornblith does not only commit himself to the idea that FA is a *viable* method of epistemological analysis, but to the stronger claim that FA provides a *complete* method, i.e., that there are no other aspects to epistemological investigation over and above identification and aggregation, as spelled out above. This brings us to (D) and the external objection that conceptual analysis is largely unnecessary. More specifically, if FA is all there is to epistemological analysis, there is no need for a thorough understanding of our epistemic concepts—as in: an understanding that goes beyond whatever semantic investigation is needed for identification—since the main component of epistemological analysis will consist in a purely empirical investigation into the epistemic phenomena picked out. In other words, if FA provides a complete methodology, we should accept (D).

But why should we take FA to provide a complete methodology? Suppose that it proves possible to successfully conduct a series of factual analyses, providing an account of what constitutes (being) *F* for *every* epistemic concept '*F*.' This would, undoubtedly, be an impressive accomplishment. But would it mark the end of epistemological investigation? Considering the picture of epistemology as the pursuit of an apt epistemic vocabulary, the question may be reformulated as follows: Would such a set of analyses necessarily yield a fully apt epistemic vocabulary? Considering what was said above in relation

---

[42] I would assume that a more straightforwardly empirical investigation also would be desirable since it would enable us to sidestep a lot of philosophical intuition mongering, in favor of straightforward empirical investigation, especially in light of the plethora of problems for traditional conceptual analysis that has been discussed in the literature. See DePaul and Ramsey (1998) for a good selection of essays discussing these problems.

to conceptual refinement, it might be tempting to answer the question in the positive, under the assumption that a more apt set of concepts simply is a more *accurate* set of concepts. In other words, the answer would be yes if the following principle can be shown to hold:

> *The AC principle*
> The pursuit of a more *apt* set of concepts reduces to that of providing a more *accurate* set of concepts.

In other words, we have established the following chain of dependency: If *the AC Principle* holds, it is reasonable to assume that FA yields not only accurate but apt concepts and, hence, provides a complete epistemological methodology. Furthermore, if FA provides a complete epistemology methodology, then it is reasonable to assume that (D) holds. However, if *the AC Principle* does *not* hold, (D) remains unwarranted since FA, thereby, might yield accurate but not necessarily apt concepts. More specifically, I will argue that

> *Problem 2*
> unless *the AC Principle* holds, FA fails to acknowledge epistemology's role in the particular kind of conceptual improvement involved in conceptual *reconstruction*.

In the following section, I will (*a*) spell out this methodological component of conceptual reconstruction, (*b*) provide two examples of cases in which the need for it indicates that increased accuracy does *not* imply increased aptness, even if we assume content externalism for the corresponding concepts, and (*c*) conclude that *the AC Principle* is an unviable epistemological assumption and (D), hence, is without warrant.

3.6. Conceptual Aptness in Science and Epistemology

A strong motivation for content externalism about natural kind terms is that it provides us with a straightforward and intuitive explanation

of disagreement in the natural sciences.[43] That is, given content externalism, successful reference is completely compatible with inaccurate concepts and theories on the part of the person referring. Hence, despite radically different (and, in some cases, incorrect) theories, Dalton, Rutherford, and modern physicists were and are essentially talking about the same phenomenon, i.e., the atom. Indeed, only if we say that can we make the further claim that they *disagree* about the latter's constitution and that contemporary theories of the atom constitute *improved* and more *accurate* theories when compared to the earlier theories of Dalton and Rutherford, in line with the overall progress of science.

Does this rather neat picture of scientific disagreement and improvement carry over to epistemology? One of the major reasons for doubting that it does is that (*a*) we have reason to believe that the majority of epistemological objects are artifactual rather than natural kinds, that (*b*) it is still to be established that content externalism provides the correct semantics for artifactual kind terms, and that (*c*) it is not obvious that explaining epistemological disagreement requires assuming *referential* continuity, rather than that our epistemological *project* is continuous with that of our epistemological predecessors (e.g., in the sense that we are all concerned with the search for an apt epistemic vocabulary). However, being able to explain disagreement is, clearly, only one motivation for content externalism and there might very well be independent reasons for wanting to defend such externalism in epistemology. For this reason, I will now consider a dialectically more robust strategy by showing that *the AC principle* does not even hold under content externalism. The strategy will first be demonstrated on an abstract level and then, in the next section, illustrated by way of two examples.

So, first consider how referents get assigned to concepts on an externalist story and let us, for dialectical purposes, assume the strongest form of externalism, i.e., *Causal Externalism*. On *Causal Externalism*, a concept gets assigned a referent by way of an initial act of

[43] See, e.g., Putnam (1975b).

baptism. In other words, the referent of a concept is fixed by the (mere) fact that the baptizer stands in a certain causal relation to it. As we have already seen, such a semantic story is perfectly compatible with successful reference in spite of considerable ignorance on part of the speaker, which leaves room for extensive conceptual refinement. Hence, it was suggested above that there is a connection between refinement and aptness.

This, furthermore, provides at least part of a rationale for the conceptual refinement of purely descriptive concepts in science, understood as concepts, the mere (or at least most central) purpose of which is to categorize without thereby providing a normative evaluation of whether something is good or bad in relation to a specified set of goals. Hence, H_2O may constitute a refinement of WATER, as far as chemistry goes, and MEAN MOLECULAR KINETIC ENERGY a refinement of TEMPERATURE, as far as kinetic theory goes. Similarly, Goldman has argued that psychology may refine the descriptive resources of epistemology, for example as it pertains to the concept BELIEF.[44]

However, to fully understand what constitutes an apt *epistemic* vocabulary, we need to keep in mind that epistemic concepts typically are *normative* concepts. As we saw in chapter 2, epistemic concepts, *qua* normative concepts, are tools by which we evaluate the conduct of fellow epistemic inquirers—that is, categorize their conduct as being an instance of knowledge, justification, rationality, etc.—but where we, merely by virtue of such a categorization, also make an explicitly normative judgment about the extent to which their conduct is good from an epistemic point of view. This suggests that the purposes of epistemic concepts may be understood in relation to the *norms* in which they figure and the *goals* that these norms are designed to meet. Furthermore, it prompts the following question: Is there any guarantee that the referent that was originally attached to an epistemic concept or term in an initial act of baptism, in fact, provides the best route to our epistemic goals?

---

[44] See Goldman (1986, pp. 199-226).

The answer is no. In analogy with what was argued in relation to *Epistemic Romanticism* above, it might, indeed, be possible to construct an argument to the effect that whatever we are referring to with our epistemic concepts does not provide a completely *useless* route to our epistemic goals, given that having true belief is an integral part of attaining many of our practical goals (and that the latter is something that we tend to do). However, it is hard to see why the referents initially determined necessarily provide the *best* paths to our epistemic goals. For this reason, epistemology has to consider the possibility that alternative referents may present a better route to our epistemic goals than our present referents. Hence, it makes sense to not only investigate the question of to what extent our concepts provide *accurate* pictures of their referents—a referent that need not provide the best route to our epistemic goal—but also whether there are any *alternative* referents that might present a better route. Such an inquiry, however, has to take into account not only the causal structure of the world (as revealed through straightforward empirical inquiry) but also the purposes of the original concepts and, in particular, the norms in which they typically figure and the goals these norms are meant to attain, since nothing short of such an investigation will enable us to specify what would constitute a better route to our goals and, consequently, a more apt concept.

3.7. When Accuracy Does Not Increase Aptness

To illustrate this point, I will now present two hypothetical scenarios in which increased accuracy does not increase aptness, since the referents in question—determined in accordance with *Causal Externalism*—do not present optimal routes to our epistemic goals. I will argue that the proper epistemological strategy in those cases is not refinement but a more substantive conceptual improvement in light of the larger context of norms and goals in which the concept figures. I will refer to such conceptual improvement as conceptual *reconstruction*, to a first approximation understood as an ameliorative activity located further out on a continuum of increasingly radical conceptual revision.

A helpful metaphor here is home improvement. When redoing, say, a kitchen, you start out with a certain pre-existing material, i.e., the kitchen that is to be redone. Let us call this kitchen K_1. The kitchen that results from the reconstruction—let us call it K_2—might look nothing like K_1. Nevertheless, K_2 will (if everything goes as planned) serve a certain set of purposes better than K_1 did. Perhaps K_2 is more spacious, has more up-to-date appliances, has a more attractive design, etc., than K_1. Indeed, it is reasonable to believe that the intention to realize those properties provided the very reason for redoing the kitchen. Similarly, conceptual reconstruction starts out with a pre-existing concept, C_1. The reconstructed concept, C_2, might, in the end, look nothing like the original concept. Still, the very point of reconstruction is that C_2 serves a set of purposes better than C_1. In the case of epistemic concepts, these purposes will be understood in relation to our epistemic goals—goals that will be spelled out in more detail below but that, to a first approximation, may be understood in terms of true belief in significant matters.

However, this raises some questions about the identity conditions for concepts. Is the concept that results from reconstruction "the same concept" as the concept that we started out with? I find this question about as puzzling (and interesting) as the question whether my kitchen remains "the same" over the course of a redecoration. Clearly, many of its properties will change—some vanish, some arise—and the redecorated kitchen will not be identical to the old one. (If it were, we would want our money back.) At the same time, it is still my kitchen and it will still serve the same purposes—indeed, it will, hopefully, serve some of the same purposes *better* than my old kitchen. Analogously, I will say that a reconstructed concept C_2 is "the same" concept as an original concept C_1 to the extent that they both figure in relation to the same set of purposes. At the same time, however, C_2 will, clearly, also be different form C_1 in the sense that it has different properties and, due to this very fact, serves the purposes in question to a greater degree.

Against the background of a distinction between conceptual refinement and reconstruction, say that we perform a factual analysis

of JUSTIFICATION and let us, for simplicity's sake, refer to this concept as *our* concept of justification.[45] Assume, furthermore, that we, in the process of identification, find that the properties by which we typically individuate degree of justification pertain to the fulfillment of epistemic duties. In fact, this would make complete sense, given what we know about the etymology of the term "justification." As noted by William Alston, the term "has been imported into epistemology from talk about voluntary action," which "explains the strong tendency to think of the justification of belief in deontological terms, in terms of being permitted to believe that *p* (not being to blame for doing so, being 'in the clear' in so believing)."[46] What does this tell us about the way that JUSTIFICATION was originally endowed with a referent? For one thing, it lends some support to the dual claim that (*a*) JUSTIFICATION—or rather "justification—was originally introduced as applying to the formation of belief, and that (*b*) it has traditionally been presupposed that we can form or refrain from forming beliefs by willed action—on pain of denying that 'ought' implies 'can'. Finally, assume that we, in the process of aggregation and empirical investigation of the phenomenon actually picked out by our concepts, find that we have no voluntary control over the formation of beliefs.[47]

---

[45] It does not matter so much for present purposes whether there is such a thing as *our* concept of justification or rather a rich multiplicity, since all that the following line of reasoning requires is that we are talking in terms of a *specific* concept—be it a widely shared one or not.

[46] Alston (1993, pp. 532 and 533, respectively). See also Plantinga (1990) and Alston (2005).

[47] The kind of voluntary control at issue here is what Alston (2005, p. 62) refers to as *basic voluntary control*. This is not the only kind of voluntary control—in fact, Alston distinguishes between three types of (decreasingly extensive) voluntary control as well as different grades of indirect voluntary *influence* (pp. 62-80). However, Alston also provides convincing arguments to the effect that, even given increased taxonomical complexity, there does not seem to be any voluntary control or influence such that it both (*a*) applies to the psychology of common epistemic inquirers, and (*b*) is sufficiently extensive to

If this turned out to be the case, we seem to have uncovered reason to believe that our concept of justification is off the mark, in that it pertains to something that we, as a matter of fact cannot have, namely epistemic duties. What would be a proper epistemological response? Two responses are available. On the first response, we reject voluntarism as a mistaken view about the way our mind works, but retain the idea that justification applies to belief-formation. This would correspond to a simple refinement and the most promising candidate for fleshing it out would probably be some form of process reliabilism.[48] However, as it stands, this response suffers from a significant problem: It takes for granted that the referent inherited from our deontological predecessors does, in fact, provide the optimal route to our epistemic goals. While this certainly cannot be ruled out, nor can it be assumed, for reasons brought out in the previous section.

This leads us to the second response, on which we engage in an inquiry best described as a continuation of aggregation, with the crucial qualification that it is preceded and guided by an investigation into the intended purpose of the original concept, in an empirical search for properties that may serve that purpose better, given a relevant set of epistemic norms and goals. For example, if such an investigation were to demonstrate that the purpose of JUSTIFICATION is to flag certain voluntary acts (previously identified as acts of belief-formation) as appropriate sources of information, and the appropriateness in question is typically understood in relation to a goal of attaining and maintaining true belief in significant matters, one possible route for empirical inquiry would be to identify a kind of voluntary act that tends to yield and support true belief, thereby providing material for a reconstructed concept. The resulting view would retain voluntarism but reject the idea that justification should apply to belief-formation. This would correspond not to a refinement but a recon-

---

warrant talking in terms of genuine duties. This matter will be attended to closely in chapter 5.

[48] See, e.g., Goldman (1986).

struction, where justification gets "re-baptized," so to speak, and JUSTIFICATION, thereby, gets assigned a new referent.

This brings us to the second scenario and the traditionally most influential candidate for such a voluntary act: *introspection*. More specifically, say that we, in the process of identification, find that we tend to determine degree of justification by reference to an introspective evaluation of reasons on part of the allegedly justified (or unjustified) subject. Let us, furthermore, assume that this particular concept originally entered into the discourse of evaluating epistemic subjects some four hundred years ago by way of Descartes' ideas about what one perceives clearly and distinctly by means of introspection.[49] What does this tell us about the way in which JUSTIFICATION was originally endowed with a referent? At the very least, it lends some support to the dual claim that (*a*) JUSTIFICATION—or rather: "justification"—was originally introduced as applying to acts of introspection, and that (*b*) it has traditionally been presupposed that such acts provide a powerful and reliable access to the grounds for our beliefs. However, suppose that we also find that, as a matter of empirical fact (say, facts uncovered by cognitive psychology), we seldom have access to the epistemic qualities of the processes by which we form beliefs and, furthermore, that the stories (consciously or unconsciously) reconstructed by us regarding the epistemic etiology of our beliefs are often quite inaccurate.[50]

If that turned out to be the case, what would be a proper epistemological response? Again, two responses are available. On the first response, we would try to identify conditions under which we *do* have reliable access to our reasons and, then, refine our concept accordingly. However, the very same research hinted at in the previous paragraph gives us reason to think that such conditions are quite hard to come by. Hence, the second response: Conduct an investigation into the purpose of our (supposedly inapt) concept, let the result of such an investigation guide further empirical aggregation of candidate

---

[49] See Descartes (1988b, p. 103; AT VII 59-60).

[50] See Wilson (2002) for some evidence to this effect.

properties that may figure in a reconstructed concept that fills the same (or close to the same) purpose, without being committed to the idea that we have a reliable introspective access to the epistemic qualities of our belief-forming processes.

In short, if an investigation into the purpose of our concept of justification were to reveal that its purpose is to flag certain voluntary acts (previously identified as acts of introspection) as appropriate sources of information, and the appropriateness is (again) typically understood in relation to a goal of attaining and maintaining true belief in significant matters, one way for empirical inquiry to proceed would be to identify an alternative kind of voluntary act (one candidate being certain acts of reasoning) that tends to yield and support true belief. An empirical aggregation preceded and guided by an investigation into the purpose of our concept would, thereby, provide material for a reconstructed concept.

These scenarios and the corresponding responses will be discussed at length in the second part of this study. In fact, I will there argue that something like the story sketched above might just provide a plausible theory about the evolution (as well as the future) of JUSTIFICATION. However, as far as present purposes go, nothing hinges on whether that turns out to be the case. Even if considered as hypothetical, the mere possibility of the above scenarios serves to highlight two points with a direct bearing on (D) and the issue of accuracy and aptness: First, there are possible cases in which merely attending to the referents of our current concepts would *not* enable us to complete the task of identifying a more apt concept, since those referents provide sub-optimal routes to our epistemic goals. For this reason, an epistemological investigation need to attend not only to the referents of our concepts but also to other properties that are not in any obvious way implicated by our current concepts but that may, nevertheless, figure in a more apt vocabulary. Second, this empirical investigation needs to be preceded and guided by an investigation into the purposes of the original concepts, providing the empirical inquiry in question with a direction in the form of an understanding of what properties to look for. For this reason, the substantially empirical

method in no way eliminates the need for an understanding of our epistemic concepts and, in particular, the particular purposes for which we employ them. On the contrary, such an understanding plays a vital role in the search for a more apt epistemic vocabulary.

Again, unless it can, somehow, be shown that scenarios like the two just considered are impossible (which is different from arguing that they are not actual), we have to leave room for the possibility of conceptual reconstruction when providing a methodological framework for epistemology. More specifically: Some instances of conceptual improvement in epistemology might indeed flow from a straightforward conceptual refinement. However, given that epistemology is concerned with explicitly *normative* concepts, some cases of improvement must take into account not only (*a*) facts about the referent but also (*b*) facts about the intended *purpose* of the concept in question, so as to guide empirical aggregation in the search for (*c*) properties that might not be implied by the original concept, nor present themselves in any straightforward way through an unconditional empirical investigation into the referent, but that might nevertheless furnish a reconstructed concept with an increased aptness, given the norms and goals that the original concept was supposed to (but failed to) meet. Hence, even if substantially empirical, a proper epistemological methodology needs to leave room for attending to our concepts, and in particular to the purpose for which we employ the epistemic concepts that we do.

More specifically, epistemological analysis can plausibly be taken to involve not only *Identification* and *Aggregation* but also the following methodological component:

> *Improvement*: To the extent that '*F*' could be refined or reconstructed so as to fulfill its purpose to a greater degree, refine or reconstruct '*F*' accordingly.

This is why epistemic concepts—*contra* the decrees of FA—do not drop out of the epistemological picture as soon as we move beyond the initial stage of delimiting a set of paradigmatic examples, and why

Kornblith's external objection to conceptual analysis misses the target. We need to investigate the purposes of the concepts at issue by incorporating an account of the norms and goals that these concepts are associated with in order to determine what would constitute a concept that served us better. As such, however, the kind of conceptual analysis called for is substantially different from the one usually assumed in the literature. Or so I will now argue.

3.8. Conclusion

The present chapter has established two conclusions. The first conclusion is that the claim that epistemology is a substantially empirical investigation is *not* contingent upon the admittedly controversial idea that all objects of epistemological investigation are natural kinds, under the plausible assumption that conceptual *refinement* plays an important role in epistemological theorizing. In fact, this can be shown to be plausible even under the dual assumption that (*a*) all objects of epistemological investigation are *artifactual* kinds, and (*b*) what I have referred to as *Strong Internalism* provides the correct semantics for artifactual kind terms. The second conclusion is that concepts are not only relevant to identification, i.e., to the fixing of a (non-exhaustive) set of (what we take to be) paradigmatic examples of the phenomenon under investigation. An insight into the purposes of our epistemic concepts is, in some cases of conceptual reconstruction, also a prerequisite for knowing how to direct the process of aggregation in the improvement of our conceptual apparatus and, hence, answering a question that ought to lie at the heart of any epistemology interested in not only describing but also improving on epistemic inquiry, namely "Given our epistemic goals, what would be a set of epistemic concepts that *served us better*?"

Chapter 4.
Constructive Analysis

In the previous chapter, we saw that epistemology is, indeed, substantially empirical, but that epistemic concepts, nevertheless, matter to epistemology in that they provide material for what I have referred to as *Identification* and *Improvement*. In the present chapter, I aim to identify what I take to be two plausible methodological candidates for playing these roles in a way compatible not only with the empirical findings that posed a problem for DCA, but also the methodological worries about exhaustiveness that surfaced in relation to PCA. In doing this, I will also elaborate further on the element of improvement identified in the previous chapter and, in particular, on how it suggests that we need to transcend the previously surveyed analyses by incorporating not only a descriptive but also an ameliorative component, in what I will refer to as *Constructive Analysis*. This is also the kind of analysis that I will implement in the second part of this study. But first, let us look closer at the aforementioned methodological components.

4.1. The Descriptive Component

In line with what was argued in the previous chapter, the first component of epistemological methodology consists in the two-fold descriptive task of (*a*) *identifying* the subject matter through an elucidation of the relevant epistemic concepts and (*b*) *aggregating* a characterization of the phenomenon referred to via these concepts. Starting with the process of identification, two comments are at place. First, the set of *relevant* epistemic concepts needs to be specified. *Contra* Goldman[1], who takes the proper starting point of epistemological analysis to be our commonsense epistemic folk concepts, and in light of Weinberg *et al.*'s study[2] and Kornblith's[3] point about the potentially beneficial influence of background theory, I will focus on the concepts not of the folk but of epistemologists, as brought out through their epistemological theories. That is, identification—the providing of the basic material for epistemological investigation in the form of a set of paradigmatic examples—should proceed by way of the categorizations made by experts, not by the folk. Construing identification thus will not only free me from the worry of assuming that there is such a thing as "our" concept of this-or-that epistemic phenomenon—over and above relevant philosophical theories—but also limit the conceptual analytic data to central epistemological texts.

There is a potential worry that needs to be addressed here, however. If there is no guarantee that the concepts, thereby, investigated—i.e., the concepts provided by epistemologists—are identical to those of the folk, what guarantee is there that our inquiry will be *relevant* to the folk, *qua* epistemic inquirers?[4] The answer to this is two-fold. First, it should be pointed out that there *is* no such guarantee—at least not as far as the analysandum goes. Since there is no guarantee that the goals of the experts coincide with those of the folk, and aptness depends on the goals relevant to the concepts used, there is no

---

[1] See Goldman (2007, 1992a).

[2] See Weinberg, Nichols, and Stich (2001).

[3] See Kornblith (2007).

[4] Thanks to Anders Tolland for calling my attention to this issue.

guarantee that the concepts of the experts will be apt in relation to the goals of the folk. However, if there is any truth to what was argued in the previous two chapters, there is not even a guarantee that the concepts of the folk will be particularly apt.

At the same time, it is reasonable to assume that experts' concepts are more accurate than the concepts of the folk, since the former are the results of a more sophisticated and methodologically sensitive interaction with epistemic phenomena. This brings us to the second point: As we saw in the previous chapter, accuracy does not imply aptness. However, accuracy facilitates *Aggregation* and *Evaluation*. The more accurate a concept we start with, the easier it will be to aggregate the relevant characteristics. And the more relevant characteristics we have aggregated, the easier it will be to evaluate whether a concept serves its purpose and its referent provides a promising route to our epistemic goals. This is why the concepts of the experts provide better basic material for analysis than the concepts of the folk; they are, in all likelihood, more accurate than those of the folk. Furthermore, and since a further task of analysis is the *Improvement* of concepts to the extent that they do not speak to the epistemic goals of the folk, we do not need an initial guarantee that the analysandum is apt, merely that it is as accurate as possible. And while starting with the concepts of the experts ensures that the analysandum is as accurate as possible, analyzing the analysandum constructively ensures that the analysans will be as apt as possible in relation to the goals of the folk.

Having mitigated that worry, we need to develop a non-traditional form of conceptual analysis that may serve the role of *Identification* but that, unlike DCA, is not committed to an implausible view of concepts. In chapter 2, we considered the extent to which PCA presented an option and I ventured to suggest that *Exhaustiveness* was not a desideratum for conceptual analysis. This point may now be put in somewhat clearer terms: The point of *Identification* is not to provide an exhaustive account of our conceptual apparatus, but just to identify a set of reasonably paradigmatic examples that may direct further empirical investigation into the characteristics of the phe-

nomenon referred to. In other words, our conceptual apparatus is, at this point, of no interest *in itself*, but only in so far as it serves to direct further investigation thus. For this purpose, we need a more manageable notion of meaning and the analysis thereof, incorporated into what we, to a first approximation, may refer to as a PCA *light*, satisfying the dual desideratum of being (*a*) based on psychological research so as to not foster an inaccurate picture of concepts, while, at the same time, being (*b*) sufficiently translucent so as to be incorporable into a useful epistemological methodology that does not yield unnecessarily convoluted outputs.

In identifying such a notion of meaning, I will start out with Hilary Putnam's notion of a *stereotype*.[5] Putnam originally introduced the notion of a stereotype in an attempt to rectify what he considered a wide-spread mistake: that of modeling natural kind terms on such words as "bachelor" that, at least on the face of it, can be characterized in terms of fairly neat conditions. What Putnam noticed was that natural kind terms do not lend themselves to such a characterization. In the case of natural kinds terms, there is no property or list of properties that captures the meaning of natural kind terms, barring such uninteresting properties as *the property of being F*. Instead, Putnam argued that the meaning of natural kind terms should be understood in terms of simplified and possibly inaccurate theories (not to be confused with any actual and usually very complex theories) about the characteristics of normal members.[6]

For instance, the stereotype for lemon would incorporate claims to the effect that typical lemons have yellow and fairly thick skin and a tart taste. Interestingly enough, these are probably also the very features we would cite if someone asked us what LEMON—or "lemon"—means. In other words, while neither (necessarily) providing an accurate picture of the phenomenon referred to, nor determining the corresponding term's reference (since we may refer to lemons with our word "lemon," even if lemons turned out to be small, blue

[5] See Putnam (1975a, 1975b).

[6] See Putnam (1975a, p. 148).

animals, who turned yellow every time humans approached), stereotypes play a non-trivial role in understanding and communication.

In particular, stereotypes can be expected to provide cognitive pathways to external phenomena. That is, while not providing anything like conditions that *determine* reference—especially not across possible worlds—stereotypes incorporate substantial information about how to identify or *fix* the reference in the actual world. For example, although it is certainly possible (although highly unlikely) that lemons might turn out to not actually have many of the properties that we typically ascribe to lemons, a conceptual stereotype incorporating such predicates as "tart," "yellow," "fairly thick skin," etc., is, undoubtedly, extremely helpful in the identification of lemons, as they tend to appear in the actual world.[7]

Now, in accordance with the fall of the Classical Theory of Concepts discussed in chapter 1, we have reason to believe that natural kinds are far from unique in not lending themselves to a characterization in terms of neat conditions. And Putnam does, indeed, consider stereotypes instances of a more general phenomenon that he refers to as "core facts":

> […] there are, in connection with almost any word (not just 'natural kind' words), certain core facts such that (1) one cannot convey the normal use of the word (to the satisfaction of native speakers) without conveying those core facts, and (2) in the case of many words and many speakers, conveying those core facts is sufficient to convey at least an approximation to the normal use.[8]

Putnam's taxonomy is slightly misleading here. Talking in terms of core *facts* (rather than in terms of stereotypes) leads one to think that we are getting at the (non-semantic) fact that *F*s are thus-and-so, rather than the (semantic) fact that "*F*" is used in a way that reveals a

---

[7] See Kripke (1980, pp. 55, 57, and 96).

[8] Putnam (1975a, p 148).

theory according to which *F*s are thus-and-so—and only the latter kind of fact leaves room for the possibility that *F*s are *not* thus-and-so. This latter possibility has to be acknowledged by any theory of meaning that is to be compatible with the highly plausible claim (discussed in the previous chapter) that, at least in the case of natural and artifactual kinds, having an accurate concept is not a requisite for successful reference. For this reason, I will simply stick to talking about stereotypes also in the general case, i.e., where Putnam uses the term core facts.

By way of conclusion: While stereotypes do not necessarily reflect accurate theories of the entities referred to, nor serve to determine the reference of the corresponding concepts, stereotypes do, however, convey the meaning of the terms expressing those concepts, by providing information that we use a certain term "*F*" so as to mean something that typically has a particular set of characteristics—characteristics that, furthermore, may serve to fix the reference in the actual world. More than this, it is reasonable to assume that, to the extent that stereotypes expose central meaning components—as in, components central to communication and understanding—they latch on to the very prototypical features mentioned in relation to PCA.

As it happens, Putnam's theory about stereotypes not only fits well with the research on prototype theory (that was just about to get underway as his work was being published), but also anticipates the so-called "theory" theory about concepts. As hinted in chapter 1 above, this theory has come to improve and supersede prototype theory by incorporating data indicating that the kind of similarity invoked by prototype theorists needs to be heavily constrained in order to predict categorization. In short, some similarities are more important than others and the evidence suggests that the kind of similarities that matter are the ones that play a central role in people's theories (in a broad sense) about the world (hence, "theory" theory). For this reason, similarity with respect to age or weight are in most cases deemed largely irrelevant in relation to whether or not something is an instance of HORSE, unlike having an udder, solid hoofs, and a mane. These kinds of observations have inspired research on

the extent to which concepts are intimately intertwined with our general theories about the world, which is the reason why the "theory" theory also has come to be known under the less misleading name *the knowledge view*.[9]

By (*a*) invoking lists of stereotypical features rather than neat conditions, (*b*) being compatible with the possibility that the component features are intimately connected with our knowledge about the world, and (*c*) opting for *Simplicity* rather than *Exhaustiveness*, stereotypes are good candidates for figuring in a refurbished conceptual analysis. To avoid confusion, however, it might be desirable to use a more neutral term. I suggest *meaning analysis*, or MA for short.[10] Consider the following:

> *Meaning Analysis (MA)*
> For any concept '*F*,' identify the stereotype for '*F*,' such that (*a*) one cannot convey the normal use of the corresponding term "*F*" without conveying that stereotype and (*b*) conveying that stereotype is sufficient to communicate at least the approximate use of "*F*."

Since the meaning is *not* conveyed via anything like a neat condition, MA is not committed to the traditional (and problematic) view of concepts associated with DCA. Instead, it has, as just noted, more in common with the prototype view, discussed in relation to PCA, or the knowledge view, with the crucial difference that it invokes a less complex notion of meaning through forsaking *Exhaustiveness* for *Simplicity* and, thereby, giving priority to simplicity of characterization through approximate accounts of meaning, over the capturing of all intuitive

---

[9] See, e.g., Murphy (2002), Carey (1999), Keil (1989), and Murphy and Medin (1985).

[10] Ernest Sosa uses the same term when he writes that some accounts of justification flow from a kind of analysis, "which leads to conclusions that no one could possibly reject without failing to understand one or another of the constitutive concepts" (BonJour and Sosa, 2003, p. 158).

judgments through strict (and excruciatingly complicated) definitions or exhaustive lists of prototypical features. Would it yield anything like analytical sentences? Not in any controversial sense. More specifically, it is reasonable to assume that MA, at most, yields analytical sentences in a very modest sense, namely in the sense that correct MAs (at least) convey approximations of the normal use of terms and that, construed thus, MAs should be obviously true to competent users of the terms. However, given the prevalence of worries regarding analyticity and analysis it might be worth considering the matter somewhat more in depth.

Although chapter 1 stressed a series of analogies between Platonic and contemporary analysis, an important disanalogy between the analyses in Plato's dialogues and modern philosophy is that the latter has been more explicitly concerned with concepts as they relate to meanings rather than to essences. However, this very move from essences to meanings is also the reason why many might consider conceptual analysis a futile inquiry, in light of W. V. O. Quine's now more than fifty year old attack on the distinction between analytic and synthetic statements in his "Two Dogmas of Empiricism."[11]

The reason is this: In the words of C. H. Langford, "the analysis […] states an appropriate relation of equivalence between the analysandum and the analysans,"[12] a relation that, within traditional analytical philosophy, came to be construed in terms of analyticity. More specifically, the very output of conceptual analysis, i.e., the resulting definition, is typically construed as an analytic statement, i.e., to a first approximation, a statement that could only come out false if the meaning of the constituent terms were to change since it, as Quine puts it, is "true by virtue of meanings and independently of fact."[13] Let us follow Paul Boghossian in referring to this particular notion of analyticity as the *metaphysical* notion of analyticity.[14]

---

[11] See Quine (1951).

[12] Langford (1942, p. 323).

[13] Quine (1951, p. 20).

[14] See Boghossian (1996).

Now, Quine criticized this notion by way of the general strategy of scrutinizing several possible ways to clarify the distinction between analytic and synthetic statements and arguing that all candidate explanantia are in as much need of explanation as the distinction itself. Hence, Quine argued, we should reject the distinction. After Quine, many philosophers have sided with Peter Strawson and Paul Grice[15] in considering Quine's rejection somewhat uncalled for—does no *clear* distinction really imply *no* distinction?—and with Hilary Putnam when he says that

> […] there is as gross a distinction between 'All bachelors are unmarried' and 'There is a book on this table' as between any two things in the world, or, at any rate, between any two linguistic expressions in the world; and no matter how long I might fail in trying to clarify the distinction, I should not be persuaded that it does not exist. In fact, I do not understand what it would mean to say that a distinction between two things *that* different does not exist.[16]

However, in understanding Quine's challenge to make sense of the distinction—a challenge that, despite the allegedly obvious character of the distinction, is still to be met—and in order to see to what extent it pertains to conceptual analysis, it is vital to acknowledge the particular employments of the distinction that Quine set out to discredit. In "Two Dogmas," the targets of Quine's critique were two positivistic projects. The first project sought to provide an explanation of necessary truth in terms of linguistic necessity, ultimately resting on conventional decisions concerning the meaning of terms. That is, every necessary truth may be stated in the form of an analytic statement, the truthmaker of which is a conventional decision about meaning, completely independent of non-semantic fact. However, as Quine pointed out, it is reasonable to assume that "truth in general

---

[15] Strawson and Grice (1956).

[16] Putnam (1975c, p. 36; emphasis in original).

depends on both language and extralinguistic fact."[17] In the somewhat clearer words of Boghossian:

> What could it possibly mean to say that the truth of a statement is fixed exclusively by its meaning and not by the facts? Isn't it in general true—indeed, isn't it in general a truism—that for any statement *S*,
>
> *S* is true iff for some *p*, *S* means that *p* and *p*?
>
> How could the mere fact that *S* means that *p* make it the case that *S* is true? Doesn't it also have to be the case that *p*?[18]

In other words, the prospect for a (purely) linguistic theory of necessary truth does not look too promising. However, Boghossian argues that it is far from clear that the second project, corresponding to the search for an account of *a priori* knowledge in metaphysically respectable terms (i.e., without reference to a faculty of rational insight), depends on this metaphysical notion of analyticity. Central to this project, he claims, is rather the idea that a statement is analytic provided that grasp of its meaning alone suffices for justified belief in its truth—a notion he refers to as the *epistemological* notion of analyticity.[19] One explanation of how such a notion could explain *a priori* knowledge and justification may be extracted from Frege's characterization of analyticity in terms of sentences transformable into logical truths by the substitution of synonyms for synonyms, because given that both synonymy and logical truths are subject to *a priori* knowledge, the same goes for analytical statements thus understood.[20]

However, even granted these admittedly substantial assumptions, the challenge remains since, at least according to Quine, there is

---

[17] See Quine (1951, p. 34).

[18] Boghossian (1996, p. 364).

[19] See Boghossian (1996, p. 363).

[20] See Frege (1953).

no way to account for synonymy without presupposing either a notion of analyticity or other notions (such as definition, intension, possibility, and contradiction) that are in as dire need of explanation as synonymy itself. And this is exactly where Quine's challenge becomes relevant to conceptual analysis. Necessary truth and *a priori* knowledge aside, it seems that the traditional practice of conceptual analysis, at the very least, presupposes a relation of *synonymy* between analysandum and analysans. However, as should be obvious from the above discussion, MA does not presuppose anything like a strong notion of synonymy or analyticity.

It might be interesting to note that Frank Jackson has suggested that Quine's own notion of a *paraphrase*, i.e., an "approximate fulfillment of likely purposes of the original sentences,"[21] is sufficient to ground the practice of conceptual analysis.[22] Furthermore, what is interesting about this suggestion is that it is clear, at least from Jackson's treatment of this notion, that conceptual analyses incorporating paraphrases involve an element of potential *reconstruction* of concepts—or a "limited change of subject,"[23] as Jackson puts it—in light of reflection. Consequently, paraphrase analysis would have to give up the *Exhaustiveness* requirement, i.e., the idea that a correct conceptual analysis admits no (genuine) intuitive counterexamples, which amounts to a move not too different from mine.

However, it needs to be stressed that giving up *Exhaustiveness* in favor of *Simplicity* has direct implications for methodology, in the sense that the dialectical framework discussed in chapter 1, where philosophical analyses are supported and refuted exclusively by reference to intuitive judgments about hypothetical situations, has to be re-evaluated. As soon as *Exhaustiveness* is given up, and, with it, the ambition to do justice to all categorization intuitions, it becomes less clear how different candidate analyses are to be assessed.

---

[21] Quine (1960, §46).

[22] See Jackson (1998, p. 45).

[23] Jackson (1998, p. 45).

At this point, however, it is vital that we see for what purpose we are at all characterizing concepts, which brings us to the second aspect of the descriptive component of epistemological methodology, namely that of aggregating a set of features characterizing not the concept at issue but the phenomenon that the concept is a concept of—which is, supposedly, what we set out to understand in the first place.[24] I will not have a lot to say about this aspect here since it should be fairly straightforwardly in line with how the corresponding kind of phenomenon is investigated in the empirical sciences. So, at this point, the analysis has to rely heavily on the kind of science relevant to the phenomenon at issue—be it psychology, physics, political science or whatever—and the fact that I will not dwell on this aspect of analysis here is not to indicate that it is in any way negligible or unimportant but rather that there is, at this point, not anything particularly philosophical about it.

In conclusion, the following sums up the descriptive component:

---

[24] I am here glossing over a subtlety to the effect that there might be cases (especially to the extent that content externalist applies) where a concept can not in any straightforward way be said to pick out *a*—as in a single and unique—phenomenon. More specifically, it might be the case that there is no fact of the matter as to *what* phenomenon a concept picks out among a series of candidates. See, e.g., Field (1973) for a discussion of such cases in relation to theory change. And while such problematic situations may *arise* within the descriptive component, I take it that they are properly *solved* within the ameliorative component, discussed below. In short, refinement and reconstruction should, in cases of referential indeterminacy, be guided by the purposes associated with the concept, motivating refinement in so far as a particular referent provides a plausible candidate given the purposes of the (inaccurate) concept, and motivating reconstruction in so far as it does not. However, these question will be discussed further in §4.2 as well as in the second part of the present study.

The Descriptive Component

Identification: For any '*F*,' identify the stereotype for the corresponding term "*F*" via Meaning Analysis, as a means to pick out a selection *Q* of what we take to be paradigmatic instances of '*F*.'

Aggregation: Against the background of an empirical investigation into the elements in *Q*, aggregate a set of characteristics that specify what actually constitutes (an) *F*.

As should be fairly obvious from the treatment of the issue in the previous chapter in relation to *Improvement*, the modest role of conceptual analysis in *Identification* or the incorporation of the explicitly empirical *Aggregation* in no way serves to make philosophy redundant. When the investigations into the (extra-mental) phenomenon in question is under way, and a set of characteristics specifying what constitutes it is being aggregated, an important philosophical task remains: that of specifying what would (if anything) constitute a more *apt* concept of the phenomenon at issue, against the background of not only what we have found out about its actual constitution but also the norms in which the corresponding concept typically figure and the goals that we desire to attain. This brings us to the second component.

4.2. The Ameliorative Component

Analytic philosophy—in so far as it has been concerned with concepts—has tended to focus on *attributional* questions regarding the correct application conditions of concepts, i.e., *that* we categorize the world in a particular way. What this approach leaves out are the further *teleological* questions pertaining to *why* we categorize the world in the way that we do. As Weinberg has argued, failure to address these questions hinders a deeper understanding of the functions of our conceptual apparatuses in the specific sense that it fails to put our concepts into context, in merely treating them as abstract conditions

on the world, rather than as potentially dynamic tools that are used for certain purposes and that, thereby, co-exist with certain norms and goals.[25]

This point was elaborated on in the previous chapter, where it was argued that the reason it is important that we focus not only on the structure of concepts but also their purposes is to leave methodological room for their improvement. It was granted that the extent to which a concept fulfills its purpose, in many situations, might just be a question of accuracy and that a mere *refinement* requires no more than an account of the ways in which the concept at issue fails to give an accurate story of its referent. At the same time, it was argued that we cannot, on strictly methodological grounds, rule out scenarios in which the referent *itself* provides a suboptimal route to our epistemic goals, in which case the inquiry called for is not a refinement but a *reconstruction* of the concept at issue. However, such a reconstruction needs to be preceded by an account of the corresponding concept's *purpose*, guiding further empirical aggregation in the search for properties the predicates of which may figure in an improved concept that better fulfills the intended purpose.

Luckily, however, the methodological (re)considerations called for in providing such accounts are not unprecedented. In his *Knowledge and the State of Nature*, Edward Craig writes:

> Let us suppose, however optimistically, that the problem of the analysis of the everyday meaning of 'know' had both been shown to exist and subsequently solved, so that agreed necessary and sufficient conditions for the ascription of knowledge were now on the table. That would be a considerable technical achievement, and no doubt a long round of hearty applause would be in order, but I hope that philosophers would not regard it as a terminus, as many writers make one feel they would. I should like it to be seen as a prolegomenon to a further inquiry: why has a concept demarcated by those condi-

[25] See Weinberg (2006).

> tions enjoyed such widespread use? There seems to be no known language in which sentences using 'know' do not find a comfortable and colloquial equivalent. The implication is that it answers to some very general needs of human life and thought, and it would surely be interesting to know which and how.[26]

According to Craig, an analysis that (unlike traditional conceptual analysis) takes into account the needs to which KNOWLEDGE answers brings out "the point of this concept, what it does for us, the role it plays in our lives."[27] More specifically, Craig suggests:

> Instead of beginning with ordinary usage, we begin with an ordinary situation. We take some prima facie plausible hypothesis about what the concept of knowledge does for us, what its role in our life might be, and then ask what a concept having that role would be like, what conditions would govern its application.[28]

Such a "practical explication," as Craig calls it, is *not* concerned with application conditions or conceptual extensions. Rather, through not only asking *what* concepts are, in fact, employed but also *why* and, thereby, for what *purposes* they are employed in ordinary situations, I take it that the practical explication may help us gain insight into the role which these concepts are supposed to play in epistemic inquiry and how they may or may not fit into different epistemic architectures. This also connects well with the idea that few (if any) epistemic terms correspond to natural rather than artifactual kinds, answering primarily to human intentions rather than categories independent of human thought. Put differently, we impose rather than discover a grid of epistemic categories on the world for particular epistemic purposes.

---

[26] Craig (1990, p. 2).

[27] Craig (1990, p. 3).

[28] Craig (1990, p. 2).

However, given that there is no guarantee that the terms within that grid play a role that serves our epistemic purposes in an optimal way, we need to leave room for the epistemic improvement of suboptimal architectures.

In light of Craig's suggestion, I take it that the following provides the first step in such an improvement:

> *Conceptual Purpose Analysis (CPA)*
> For any '*F*,' identify a set of *paradigm situations* in which '*F*' is utilized, and scrutinize these situations so as to identify the norms in which '*F*' typically figures and the goals these norms typically are meant to attain, in order to specify the *purpose* for which we employ '*F*.'

In line with what was argued in the previous chapter, the point of conducting such an analysis is to evaluate the extent to which the concept in question—as unveiled through MA—meshes with the norms in which it figures and facilitates the attainment of the goals at which these norms are aiming.[29] In so far as it does not, something has to go. I will assume that the components of architectures make up a natural hierarchy, corresponding to a decreasing appropriateness of revision. More specifically, I will assume that the kind of evaluation at issue should take the set of goals as a fixed starting-point, not open to revision, and that the proper objects of revision are concepts, an alteration of which, indirectly, gives rise to an alteration of the corresponding norms. This assumption is to mirror the plausible idea that, in light of a failure to attain our goals, a re-evaluation of the goals rather than the tools devised in attempting to attain them should be the last resort—not the first.

---

[29] Within epistemological analysis, this component bears some similarity to the more general project within the study of reason that Samuels, Stich, and Faucher (2004) refer to as *the evaluative project*, concerned with the extent to which human reasoning accords with appropriate standards, given some criterion of what constitutes good reasoning.

So, what are the goals relevant to the evaluation of epistemic concepts? In answering this question, I would like to make a distinction between *domain general* and *concept specific* goals. The former kind of goals pertains to concepts just by virtue of the general domain in which the concept is employed. For example, there are certain epistemic goals relevant to concepts just by virtue of the fact that they are *epistemic* concepts and, in particular, certain goals pertaining to truth—or so I will argue.[30] Over and above these goals, however, there are also certain concept specific goals that are more or less directly related to these general goals, and that serve to specify the more fine-grained roles that different concepts play within epistemic architectures at large. Consequently, while both knowledge, justification, and understanding may, in part, be specifiable with reference to the epistemic goal of truth, they each play different (although potentially interrelated) roles by virtue of their respective concept specific goals—say, pertaining to warrant in the case of knowledge, reliability in the case of justification, and quality of explanation in the case of understanding—that are more or less directly related the domain general epistemic goal of attaining truth.

Since this chapter concerns general epistemological methodology, it will focus exclusively on domain general goals. A treatment of concept specific goals will have to wait until the second part of this study (§7.2, to be exact), which will be concerned with the particular concept of epistemic justification. In accordance with the taxonomy introduced above in relation to epistemic normativity, the methodology to be defended will be working with sets of epistemic desiderata, specifying that which constitutes or is a means to a epistemic goals. Furthermore, under the assumption that goals and their satisfaction give rise to values, we may say that a desideratum of an endeavor is a specification of something that is *valuable* in relation to the goals of

---

[30] See Stich (1991) for a series of arguments again the idea that epistemic goals should be construed in terms of truth and Kornblith (2002) for a critical discussion and, to my mind, a rebuttal.

that endeavor.[31] Hence, all properties that *satisfy* desiderata are valuable in relation to the inquiry to which those desiderata pertain. As for the source of particularly epistemic desiderata, I will take it that a pivotal point has to be the cognitive goal pertaining to the desire that "our beliefs correctly and accurately depict the world," to quote Laurence BonJour.[32] This goal defines what Alston has referred to as "the epistemic point of view"[33] from which epistemic evaluation is undertaken, and may to a first approximation be characterized in terms of the *synchronic* epistemic desideratum that

(D1) it is *now* the case that our beliefs are true.

Against the background of this desideratum, truth is, clearly, of epistemic value. Furthermore, everything that is *conducive* to truth is of (instrumental) epistemic value.

However, (D1) is not the sole desideratum of epistemic inquiry. For one thing, any strict skeptic would satisfy (D1) by not believing anything. One way to counteract this is to acknowledge the *diachronic* aspect of our cognitive goal, pertaining to the way in which we want our belief set to evolve dynamically—an aspect arising out of the fact that epistemic inquiry is an inquiry stretching over time. This aspect may be characterized in terms of the diachronic epistemic desideratum that

(D2) we *attain* and *maintain* true beliefs.[34]

---

[31] This is certainly not to say that goals are the only source of value. The way I will talk about value in the following is completely compatible with there being values completely independent of goals and the meeting thereof.

[32] BonJour (1985, pp. 7-8). See also Moser (1985, p. 4) and Alston (2005, p. 30).

[33] Alston (1989, p. 83).

[34] Cf. Weinberg (2006, p. 36) on *diachronic reliability* as an epistemic desideratum.

In other words, we do not only want our beliefs to be true *now*—a desire brought out through the synchronic desideratum (D1)—but also for it to be the case that our belief set expands and contracts over time so as to (*a*) *attain* true belief about that which we currently have false or no beliefs about, and (*b*) *maintain* true beliefs about that which we currently have true beliefs about. Differently put, we want it to be the case that our belief set fluctuates in the particular sense of *tracking the truth*.[35]

Three things should be noted about these desiderata. First, I do not wish to claim that they are the *only* epistemic desiderata. Nevertheless, I take them to be sufficiently central to epistemic inquiry to provide a not too distorted basis of evaluation. This centrality is also to be expected if we consider why we not only *do* but also *should* (as in: have reason to) value truth. To see this, first consider the connection between true belief and successful action, spelled out by Goldman thus:

> The pragmatic utility of true belief is best seen by focusing on a certain subclass of beliefs, viz., beliefs about one's own *plans of action*. Clearly, true beliefs about which courses of action would accomplish one's ends will help secure these ends better than false beliefs. Let proposition *P* = "Plan *N* will accomplish my ends" and proposition *P'* = "Plan *N'* will accomplish my ends". If *P* is true and *P'* is false, I am best off believing the former and not believing the latter. My belief will guide my choice of a plan, and belief in the true proposition (but not the

[35] The intended notion of truth tracking is *not* identical to the one utilized by Nozick (1981) in his counter-factual condition on knowledge (stating that if one knows that *p*, then (*a*) if it were not the case that *p*, one would not believe that *p*, and (*b*) if it were the case that *p*, one would believe that *p*). Truth tracking, as I understand it here, is not a modal but a *probabilistic* notion, applying to belief sets in so far as they are likely to fluctuate with changes in the world. Construed thus, the set of belief sets that track the truth in Nozick's sense is a subset of the belief sets that track the truth in my sense.

> false one) will lead me to choose a plan that *will* accomplish my ends.[36]

In other words, accurate beliefs about plans for action increase the chances of accomplishing one's ends. If so, however, we all seem to have reason to value truth. Hence, Kornblith:

> It seems that someone who cares about acting in a way that furthers the things he cares about, and that includes all of us, has pragmatic reasons to favor a cognitive system that is effective in generating truths, whether he otherwise cares about the truth or not.[37]

The reason is that, in the course of pursuing what we desire, we need to make evaluations of alternative actions. And whatever it is that we happen to desire, we have reasons to want such evaluations to be done accurately. Alternatively put, we have reasons to want whatever cognitive system we employ in our pursuit of that which we happen to value, to be one that, at the very least, tends to generate truths.[38]

This relation between truth and the satisfaction of desires brings us to the second point: As they stand, (D1) and (D2) figure at a level of abstraction that makes them unfit for the prediction of actual human behavior. This is because (D1) and (D2) are purely *epistemic* desiderata, while we, as human beings, are not purely epistemic beings. In other words, (D1) and (D2) capture but an *aspect* of human inquiry, namely an *epistemic* aspect and this aspect may be neutralized in more naturalistic and less abstract settings. However, as Marian David points out:

> We have various goals. In some cases different goals come into conflict and 'loses out.' […] This does not show that we

---

[36] Goldman (1992b, p. 164).

[37] Kornblith (2002, p. 156).

[38] See Alston (2005, p. 31) for a similar argument.

> don't have the beaten or neutralized goal; it merely shows that the goal is not an absolute one.[39]

Still, in understanding the interplay between different aspects of human inquiry, and between epistemic and non-epistemic inquiry in particular, we need to take into account that our reasons for truth are intertwined with varying sets of desires, some of which are not purely epistemic in nature. More specifically, "our desire for truth is largely coordinate with our desire for answers to our various questions," to quote Ernest Sosa.[40] In other words, we do not value truth *as such*, nor do we necessarily have *reason* to do so. If we did, "we could not better spend our time than by memorizing telephone directories,"[41] as William Alston points out. This point generalizes to all purely epistemic desiderata and calls for a qualification of (D1) and (D2).

Furthermore, this particular point has methodological implications. Needless to say, evaluating our epistemic concepts along an exclusively epistemic dimension via purely epistemic desiderata enables us to pinpoint and characterize a very particular aspect of human life concerned with the pursuit of truth. However, as we just argued, this is just *one* aspect and an aspect that, in real life settings, is intimately intertwined with non-epistemic desires and desiderata. Hence, an epistemology that focused exclusively on this epistemic aspect would do a terribly bad job of addressing the epistemological questions pertaining to the guidance or improvement of actual epistemic inquiry for the simple reason that it would not take into account the naturalistic settings in which such pursuits of truth actually take place. Consequently, it would fail miserably on the third and *ameliorative* task, pertaining to that it is reasonable to ask of the epistemologist to improve on relevant epistemic concepts to make for a better fit with the goals and desiderata found in not just *any* epistemic architectures but the very ones that figure in actual truth seeking situations.

---

[39] David (2005, p. 299).

[40] Sosa (2003, p. 157).

[41] Alston (2005, p. 32).

Following Kitcher as well as Michael Bishop and J. D. Trout, I will approach this question in terms of *significance*.[42] To a first approximation: Since our interest in truth is contingent upon our wants and desires, and the latter may very well vary between persons and cultures, some questions, and the truths they pertain to, are deemed more important than others, even though they may all be equally significant from a narrowly epistemic (and glaringly abstract) perspective. In discussing the particular epistemic inquiry of science, Kitcher makes an illustrative analogy with maps.[43] The accuracy of maps is, clearly, dependent on (although not determined by) the interests of whoever makes use of the maps. By the same token, Kitcher claims that "the aim of the sciences is to address the issues that are significant for people at a particular stage in the evolution of human culture" and that scientific languages "are fashioned to draw those distinctions that are most helpful in carrying out the lines of investigations those people want to pursue."[44] Kitcher extends this analogy a step further:

> Like maps, scientific theories and hypotheses must be true or accurate (or, at least, approximately true or roughly accurate) to be good. But there is more to goodness in both instances. Beyond the necessary condition is a requirement of significance that cannot be understood in terms of some projected ideal—completed science, a Theory of Everything, or an ideal atlas. Recognizing that the ideal atlas is a myth, I hope to have provoked concerns about the analogue for inquiry generally. A rival vision proposes that what counts as significant science must be understood in the context of a particular group with particular practical interests and with a particular history.[45, 46]

---

[42] See Kitcher (2001) and Bishop and Trout (2005).

[43] Cf. Giere (1999).

[44] Kitcher (2001, p. 59).

[45] Kitcher (2001, p. 61).

Kitcher proceeds to give an account of the source of significance through a discussion of (*a*) the mechanisms by which we deem certain items—such as questions, answers, claims, hypotheses, apparatuses, methods, etc.—as significant and (*b*) how significance, furthermore, may flow via the interconnections of such items. He suggests that the best way to illustrate the workings of these mechanisms is through *significance graphs*, where the items are represented as nodes connected by arrows, signifying the inheritance of significance.[47] Such significance graphs thus provide a sketch of the interdependence of different items for a specified field of inquiry with regard to their significance. However, an important question remains: Is there any room for not only talking about what people *do* find significant—as brought out through such graphs—but also what they *should* find significant?

To answer this question we need to look into not only how desires and interests serve to mark certain items as significant, but also how certain items may *deserve* that mark better than others. Focusing on the significance of problems and questions, the latter issue has been addressed lately by Bishop and Trout in the form of what they

---

[46] Similarly, J. L. Borges (1999) writes, in his one-paragraph short story titled "Of the Exactitude of Science," about a fictional empire in which "the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography."

[47] See Kitcher (2001, p. 78).

refer to as "the thin-thick problem."[48] This problem arises out of the challenge to identify a notion of significance that is *thick* enough to exclude the "significance" resulting from powerful but clearly deviant desires to answer certain questions (say, pertaining to the measurement of one's thumbnail every five seconds), yet *thin* enough to allow for substantial interpersonal differences. Focusing on the significance of problems, they suggest that

> […] the *significance* of a problem for *S* is a function of the weight of the objective reasons *S* has for devoting resources to solving that problem.[49]

As for the specific nature of such objective reasons, Bishop and Trout claim that some reasons (for example the ones arising out of basic moral obligations) simply are universal, while others (such as those arising out of one's social or professional obligations), clearly, may vary from person to person. The ultimate determining factor of significance, however, is conduciveness to *human well-being*, where both well-being and the conditions that are conducive to it are open to empirical investigation. More specifically, I interpret them as claiming that

> the objective *weight* of *S*'s epistemic reasons is, ultimately, a function of the extent to which solving the problem is conducive to *S*'s well-being.[50]

Before drawing out the implications for the discussion at hand, let us recapitulate: We have established that we value truth (which is not to say that we value all truths equally) and, moreover, have *reason* to do so given that we have any desires at all, since true belief happens to be

---

[48] See Bishop and Trout (2005).

[49] Bishop and Trout (2005, p. 95; emphasis removed from all words but one).

[50] This is my formulation/reconstruction, not a quote. See Bishop and Trout (2005, p. 99).

a good means to satisfy our desires and needs. This, furthermore, seems to be mirrored in the fact that we value truths to the extent that they pertain to questions we want to answer, issues we want to address, etc. And since the latter may very well vary between persons, cultures, etc., only a particular subset of truths (and, consequently, questions, answers, issues, etc.) is deemed significant—i.e., worthy of our pursuit given limited time and resources. Given that some basic desires and needs are constant across persons and cultures, however, we have reason to expect that there will be an overlap between such sets, although there will, most likely, also be some quite substantial inter-personal differences, since some desires and needs, clearly, arise out of—and, hence, are dependent upon—highly contextual factors (such as who you are, where you are, what you are doing, etc.).

This is exactly what gave rise to the thin-thick problem and I am sympathetic to Bishop and Trout's attempt to tackle it by recourse to objective reasons. I am also sympathetic to their idea that a philosophical theory of significance, when considered in isolation, must be incomplete since it needs to incorporate direct empirical investigation of what people desire and what is conducive to satisfying these desires, as well as the extent to which people's predictions of the impact of future events on the satisfaction of their desires are accurate.[51]

However, I am more skeptical as to their attempt to construct a normative notion of significance by invoking human well-being. As Bishop and Trout rightly notes, there is quite strong evidence indicating that what ultimately contributes to the well-being and happiness of people are such things as health, deep social attachments, personal security, and the pursuit of meaningful projects,[52] and, furthermore, that certain economic, social and political institutions have a systematic and adjustable effect on the creation of happy people.[53] And this is relevant to epistemic inquiry and epistemology to the extent that *any*

---

[51] See, e.g., Gilbert *et al.* (2002).

[52] See, e.g., Diener and Seligman (2002).

[53] See Diener (2000), Diener and Oishi 2000), and Frey and Stutzer (2002).

intellectual inquiry reasonably should strive to facilitate—or at the very least not *counteract*—the creation of conditions under which well-being is maximized.

My skepticism about the posited connection between significance and human well-being can be summed up in two points: (*a*) Considerations regarding human well-being does not seem to track very well the issues we should consider significant, and (*b*) there is a more straightforward way to solve the problem that does not rob perfectly notable issues of their significance. As for the first point: The truths that we should consider significant and the truths the knowing of which is conducive to human well-being are not as intimately connected as Bishop and Trout seem to assume. In particular, I take it that there are substantial sets of truths, questions, problems, etc., that are, at best, connected to the promotion of human well-being in a very indirect way, but that still seem to be perfectly significant. Several good examples may be found among the truths, questions, and problems pertaining to certain foundational issues in mathematics, natural science, and metaphysics. Is it always possible to find, somewhere in the infinity of integers, a progression of any length of equally spaced prime numbers? Are there three, four or eleven dimensions to our universe? Are there non-physical substances? Does space and time consist of gunky or continuous, extensionless points? It is not clear to me that any of these questions may be related to questions of human-well being, nor why that ought to imply that they should not be considered significant.

This brings us to my second point: There is an alternative way to solve the thin-thick problem that does not rob what seem to be perfectly notable problems and questions of their significance. Consider the following slightly altered version of Bishop and Trout's suggestion:

> The *epistemic significance* of a problem for *S* is a function of the weight of the objective *epistemic* reasons *S* has for devoting resources to solving that problem, where the objective *weight* of *S*'s epistemic reasons is a function of the extent to which solv-

> ing the problem is conducive to addressing *S*'s interests and satisfying her desires.

The fact that the objective weights are tied to and vary with *S*'s contingent interests and desires serves to make our notion of significance thin enough to both allow for inter-personal differences and acknowledge the fact that epistemic significance is intimately connected to practical ends and needs as well as social setting. However, it is, as it stands, not thick enough to rule out clearly deviant interests and desires as fully legitimate sources of significance. Unlike Bishop and Trout, however, I will not thicken the suggestion by reference to what people ought or ought not to do in relation to the production of conditions conducive to human well-being, but rather with reference to what people, as a matter of empirical fact, tend to do. Hence, Kitcher, when discussing how significance enters into significance graphs in the first place, or what he calls the "ultimate source of significance":

> Partly as the result of our having the capacities we do, partly because of the cultures in which we develop, some aspects of nature strike us as particularly salient or surprising. In consequence we pose broad questions, and epistemic significance flows into the sciences from these. [54]

Kitcher continues:

> Human beings vary, of course, with respect to the ways in which they express surprise and curiosity. Some are disposed to ask more, others less. Typically, we respond to the diversity with tolerance, explaining some of the variation in terms of difference in cultural or educational context. But tolerance has its limits, and we do count some of our fellows as pathological, either because they obsess about trifles or because they are

[54] Kitcher (2001, p. 81).

> completely dull. In claiming that the sciences ultimately obtain their epistemic significance from the broad questions that express natural human curiosity, I am drawing on this practice of limited tolerance, on our conception of "healthy curiosity" and the commonplace thought that most of us, given minimal explanation, would find interesting the global questions that stand at the peripheries of significance graphs.[55]

So, in short, the claim is that most people are subject to a healthy curiosity and, hence, desire to know certain truths (address certain questions, solve certain problems, etc.) rather than others. Furthermore, this picture of man as subject to a healthy curiosity suggests that there, as a matter of empirical fact, is a non-negligible overlap when it comes to what people take to be significant, and that certain truths are (at least at a fairly basic level) quite consistently deemed significant. This serves as a contingent restriction not upon what people *can* consider significant (which, as far as the definition above is concerned, may vary without constraint), but upon what people *actually do* consider epistemically significant, as a result of the healthy curiosity that most of us are subject to and that drives us towards certain questions, truths and issues rather than others—be it within ornithology, public policy, or foundational natural science.

This, however, does not imply that there is no room for normative judgments on the part of the epistemologist as to the extent to which subjects assign *proper* weights to their reasons for deeming something significant. Since such weights, on this account, are directly tied to interests and desires, the extent to which solving certain problems *actually* are conducive to the addressing of interests or satisfaction of desires is perfectly open to empirical investigation. At this level of analysis, however, the objective is not to deliver hands-on prescriptive advice but rather to provide a framework in which such prescription can be made in a coherent and productive way, all in the service of facilitating our various pursuits. This brings us back to where we started: Apart from describing and evaluating

[55] Kitcher (2001, p. 81).

where we started: Apart from describing and evaluating epistemic concepts from an abstract epistemic perspective, epistemology should also involve an ameliorative component, corresponding to the task of actively improving epistemic concepts as they figure in epistemic norms, in an effort to promote the satisfaction of epistemic desiderata pertaining to the naturalistic settings of real-world truth-seeking inquiry. Furthermore, and if there is any truth to what has just been argued, the following reformulated desiderata should play a pivotal role in this potentially revisionary task:

(D1*) As far as significant truths go, it is *now* the case that our beliefs are true.

(D2*) We *attain* and *maintain* true beliefs about significant matters.

To recapitulate, my disagreement with Bishop and Trout regarding the proper way to construe significance is over whether significance ultimately should be tied to human well-being or to whatever we happen to find interesting, as a result of our contingent wants and desires. Against the background of Kitcher's notion of healthy curiosity, and the underlying empirical assumption that there is a substantial overlap in wants, desires and, as a result, the issues that we tend to find significant, I opted for the latter alternative. The virtue of this approach is that it enables us to solve the thin-thick problem by allowing for interpersonal differences as well as normative evaluation and rectification to the extent to which something, as a matter of fact, is not conducive to satisfying our interests and desires, without having to rob what seems to be perfectly notable issues—e.g., foundational issues in mathematics, natural science, and metaphysics—of their significance just because they may turn out to be unrelated to issues of human well-being.

Having thus spelled out the notions of CPA, significance, as well as a relevant set of epistemic desiderata, the following sums up the ameliorative component of epistemological methodology:

The Ameliorative Component

Evaluation: For any '*F*,' identify the purpose of '*F*' via CPA, including the specific role that '*F*' is supposed to play in relation to the domain general desiderata (D1*) and (D2*). Then, evaluate the extent to which '*F*'—as unveiled through Meaning Analysis—could be refined or reconstructed so as to fulfill its purpose to a greater degree.

Improvement: To the extent that '*F*' could be refined or reconstructed so as to fulfill its purpose to a greater degree, refine or reconstruct '*F*' accordingly.

Due to its commitment to such an emendation-oriented approach, I will refer to the kind of analysis worked out here—that is, the descriptive and ameliorative components taken together—as *constructive analysis*. The term is chosen because of the dual fact that constructive analysis is designed to (*a*) enable not only the refinement but also the *reconstruction* of epistemic concepts to the extent that they do not mesh with the relevant desiderata, as well as to, thereby, (*b*) *serve a useful purpose* for actual epistemic inquiry in naturalistic settings.[56]

4.3. Conclusion

This chapter has argued that epistemological methodology is best viewed as consisting of two components: a descriptive and an ameliorative. The former comprises two sub-components and resembles DCA in that it starts out with conceptual elucidation, and PCA in not working with neat conditions, but differs from both in not being aimed at providing exhaustive accounts of our epistemic concepts. Instead, the conceptual elucidation merely serves to facilitates the *identification* of a set of paradigmatic instances through what I called

[56] In these respects, constructive analysis bears some similarity to Rudolph Carnap's notion of *explication*. See Carnap (1950).

Meaning Analysis, providing the material for the second subcomponent: an empirical *aggregation* of the properties characterizing the referent.

As we saw in chapter 3, this can not be all there is to epistemological analysis, which brings us to the second and ameliorative component, concerned with an *evaluation* of the extent to which our conceptual apparatus serves the purposes implicit in the norms and goals in relation to which it figures, and an *improvement* of it to the extent that it does not. I ventured to suggest that what I call Conceptual Purpose Analysis would do a better job than traditional conceptual analysis here, by focusing not so much on the exact make up of concepts as on the situations in which those concepts are used and, in particular, what such situations and use reveal about the purposes of our concepts and how they could be improved to serve their purposes better.[57]

Taken together, I referred to these two methodological components—the descriptive and the ameliorative—as *constructive analysis*, a term alluding to the idea that such analysis will not only enable us to refine as well as (re)construct new concepts that serve us better, but also serve a useful purpose for actual epistemic inquiry. Still, the proof is in the pudding. So, in order to fill in the gaps of this still quite abstract framework, as well as demonstrate it in action, I will now turn to a constructive analysis of an epistemic concept that has received a lot of attention within epistemology of late: *epistemic justification*.

[57] See Weinberg (2006) and Craig (1990).

Part II.

Epistemic Justification—
A Constructive Analysis

Chapter 5.
Justification and Epistemic Duties

We now have a framework for constructive analysis and the task of this part of the study is to put this framework to work. Starting with this chapter, I will implement it in an analysis of *epistemic justification*—a phenomenon that has been subject to a lot of epistemological scrutiny ever since the publication of Gettier's critique of the classical tripartite analysis of knowledge.[1] However, for reasons discussed in chapter 3, I will not assume that there is such a thing as "our" concept of justification. Instead, I will focus on the rich variety of analyses that have been delivered over the years and, in an attempt to identify our subject matter, start by pulling out two inter-related themes. The first one, which will be discussed in this chapter, pertains to *deontologism*, i.e., the idea that to be justified consists in having fulfilled one's epistemic duties and obligations. The second one—a view I will refer to as *introspection-based access internalism* and that will be discussed in the next chapter—pertains to the idea that to be justified consists in having paid due attention to one's reasons by way of introspection.

[1] See Gettier (1963).

When turning to the process of aggregation, i.e., the spelling out of the actual ontological commitments and empirical consequences of these two views, both suggestions will be shown to be problematic in rendering the corresponding concept of justification largely ineffective in the pursuit of our epistemic goals. More specifically, deontologism fails due to an implausible commitment to the idea that we have voluntary control over the formation of our beliefs, while introspection-based access internalism fails because of an overtly optimistic view of our introspective capabilities.

5.1. Deontologism and Epistemological Guidance

Epistemological deontologism is the idea that epistemic terms are best understood in terms of epistemic duties and obligations (I will use two interchangeably). As it happens, Alvin Plantinga has persuasively argued that epistemologists have traditionally thought about justification in deontological terms—at least if we let the tradition be defined by such philosophers as John Locke and René Descartes.[2] Hence, we find Locke in *An Essay Concerning Human Understanding* talking about our "duty as a rational creature" and saying that

> He that believes, without having any reason for believing, may be in love with his own fancies; but neither seeks truth as he ought, nor pays the obedience due his maker, who would have him use those discerning faculties he has given him, to keep him out of mistake and error. He that does not this to the best of his power, however he sometimes lights on truth, is in the right but by chance; and I know not whether the luckiness of the accident will excuse the irregularity of his proceeding. This at least is certain, that he must be accountable for whatever mistakes he runs into: whereas *he that makes use of the light and faculties God has given him, and seeks sincerely to discover truth, by those helps and abilities he has, may have this satisfaction in doing his duty as*

[2] See Plantinga (1990). See also Alston (2005).

> *a rational creature, that though he should miss truth, he will not miss the reward of it.* For he governs his assent right, and places it as he should, who in any case or matter whatsoever, believes or disbelieves, according as reason directs him. He that does otherwise, transgresses against his own light, and misuses those faculties, which were given him […][3]

In other words, a sincere pursuer of truth, who makes responsible use of the epistemic faculties that God has endowed her with, is not to blame if she does not get at the truth. A person who believes without reasons, however, and that does not pursue truth in the responsible way God intended her to, may come upon the truth by accident ever so often, but is nevertheless accountable for her epistemically indolent ways.

No more than 50 years back in time from Locke's publication of *Essay*, we find another founding father of western epistemology, René Descartes, and his *Meditations on First Philosophy*. Descartes claims that God has bestowed us with a free will, but since "the scope of the will is wider than that of the intellect"[4] it is possible for us, as finite human beings, to go intellectually astray. However, God is not to blame for this. It is, as Descartes puts it, "undoubtedly an imperfection in me to misuse [my] freedom and make judgments about matters which I do not fully understand."[5] Although Descartes is not as explicit as Locke about the element of duty here, he writes:

> If […] I simply refrain from making a judgement in cases where I do not perceive the truth with sufficient clarity and distinctness, then it is clear that I am behaving correctly and avoiding error. But if in such cases I either affirm or deny, then I am not using my free will correctly. If I go for the alternative which is false, then obviously I shall be in error; if I take

[3] Locke (1996, book IV, chapter xvii, 24; my emphasis).

[4] Descartes (1988b, p. 102; AT VII 58).

[5] Descartes (1988b, p. 104; AT VII 61).

> the other side, then it is by pure chance that I arrive at the truth, and I shall still be at fault [*alternative translation*: and I do not escape the blame of misusing my freedom] since it is clear by the natural light that the perception of the intellect should always precede the determination of the will. In this incorrect use of free will may be found the privation which constitutes the essence of error.[6]

More recently, we find Roderick Chisholm—one of the 20th century's most prominent epistemologists—defining justification in terms of the relation "more reasonable than," and suggesting that

> We may assume that every person is subject to a purely intellectual requirement—that of trying his best to bring it about that for every proposition *h* that he considers, he accepts *h* if and only if *h* is true. One might say that this is the person's responsibility or duty *qua* intellectual being […] One way, then, of re-expressing the locution '*p* is more reasonable than *q* for *S* at *t*' [where '*p*' and '*q*' range over doxastic attitudes, not propositions] is to say this: '*S* is so situated at *t* that his intellectual requirement, his responsibility as an intellectual being, is better fulfilled by *p* than by *q*."[7]

So, the common denominator in the works of Locke, Descartes, and Chisholm is an idea that justification pertains to the fulfillment of epistemic duties—an idea that we, in the language of Meaning Analysis, may sum up in the following stereotype:

> *Epistemological Deontologism*
> That *S* is epistemically justified in believing that *p* means that *S* has fulfilled her epistemic duties in forming her belief that *p*.

---

[6] Descartes (1988b, p. 103; AT VII 59-60). Alternative translation from Descartes (1955, p. 176)

[7] Chisholm (1977, p. 14).

Henceforth, I will often leave the qualifier "epistemic" implicit when talking about epistemic duties.

Now, as has been noted by Alvin Goldman, this conception of justification has, historically, been paired with an idea to the effect that "one central aim of epistemology is to guide or direct our intellectual conduct."[8] Or, as Kornblith has put the point more recently:

> […] a historically important motivation for engaging in epistemological theorizing, and, indeed, more than this, a philosophically important motivation for engaging in epistemological theorizing, is the idea that an adequate epistemological theory would guide the concerned epistemic agent in the conduct of inquiry. We are interested in epistemology precisely because we desire to improve our epistemic performance; an adequate epistemology ought to tell us how to achieve such improvement.[9]

More specifically, we are dealing with what I will refer to as *the Guidance Conception of Epistemology*, an idea that may be summed up in the following thesis:

> *The Guidance Conception of Epistemology (GC)*
> The main epistemological desideratum is to *guide* epistemic inquiry, in providing means for epistemic inquirers to fulfill their desiderata.

As above, a desideratum is a specification of conditions constitutive or conducive to a certain goal (or set of goals). Hence, an *epistemological* desideratum is a specification of conditions under which a certain *epistemological* goal is met. On the guidance conception, one central and important goal for epistemology is to guide epistemic inquirers—an

[8] Goldman (1999, p. 272).

[9] Kornblith (2001, pp. 242-243).

approach in the tradition of Descartes and his *Rules for the Direction of our Native Intelligence*.[10] Hence, anything constitutive or conducive to that goal (such as, say, formulating plausible action-guiding norms for epistemic agents to follow) is an epistemological desideratum.

Not surprisingly, a commitment to GC has implications for the kind of epistemological work that is taken to be worthwhile, perhaps even for what is taken to be legitimate instances of epistemology. Returning to the issue of justification, we find Mark Kaplan arguing against "a vision of a pure theory of justification, separated from all that would make it methodologically potent, cleansed of the concerns proper to the realm of the ordinary."[11] He continues:

> The problem is that a theory of justification thus purified, a theory of justification deprived of any role in methodology or the conduct of inquiry and criticism, is a theory that divorces epistemology from the very practices that furnish it with its only source of intuitive constraint. It is epistemology on holiday.[12]

In light of what was argued in the previous part of this study, Kaplan's talk about "intuitive constraint" is potentially misleading. It is important that we separate the issue of the role of categorization intuitions in the construction of philosophical theories of justification from that of grounding epistemology in actual epistemic practice. We are here concerned with the latter. According to Kaplan, epistemology removed from regulative concerns reduces to "an exercise in pure stipulation."[13] Even though he does not explicitly say that "epistemology on holiday" is illegitimate, it is fairly clear that he considers it to be less legitimate than the kind of epistemology—according to Kaplan, stemming from "a noble tradition […] exemplified in the

[10] See Descartes (1988a; AT X 359-435).

[11] Kaplan (1991, p. 154).

[12] Kaplan (1991, p. 154).

[13] Kaplan (1991, p. 148).

seventeenth century by Descartes [and] in the first half of our own century by the authors of logical empiricism"—that remains at work and "seeks to clarify, to criticize, to improve the conduct of inquiry and criticism," and constructs theories that "confront deep methodological issues and evaluate the way in which inquiry is properly to be conducted."[14]

This idea of epistemology as an essentially guidance directed endeavor may be linked up with Deontologism in such a way that, if Deontologism holds and justification is (merely) a matter of fulfilling certain epistemic duties, then a central desideratum for epistemic inquiry is to answer the following question:

> (Q1) How do I, *qua* epistemic agent, act in accordance with my epistemic duties?

However, in order to answer (Q1), we first need to know what our epistemic duties *are*. Consequently, attempts to spell what our epistemic duties consist in—be it to believe in accordance with one's evidence, God's will, or whatever the relevant factor turns out to be—will serve to guide our epistemic pursuits. Hence, the connection between Deontologism and GC.

Before at all attempting to spell out what these duties consist in, however, we need to get clearer on what it is to have a duty in the first place. As will be demonstrated below, the duties of Deontologism is best understood in terms of the following 'ought' implies 'can' principle:

> (OC) You can only have an epistemic duty to believe something if you have the ability to chose voluntarily whether or not to form a belief in accordance with that duty.

---

[14] Kaplan (1991, p. 154).

As I will now show (while not claiming that I am the first to do so), Deontologism construed along the lines of (OC), or what I will refer to as *Strong Deontologism*, constitutes an untenable position. In §5.4 I will also discuss and reject an attempt to defend what I will refer to as *Weak Deontologism*—a position that tries to evade the problems of *Strong Deontologism* by relinquishing (OC). I will thus reject both takes on Deontologism and, in want of a more plausible construal, conclude that justification should not be understood in terms of epistemic duties.

5.2. Strong Deontologism and the Problem of Voluntarism

What is fundamentally puzzling about understanding justification along deontological lines is that such talk seems to imply that what is being evaluated is subject to voluntary control. Hence, Kornblith:

> If I can't be heard in the back of the room when I'm presenting a paper at a conference, then I may be criticized for not having spoken louder: I should have spoken louder, it will be said, and rightly so. I am criticizable here, it seems, because I could have spoken louder but didn't; how loud I speak is subject to my voluntary control. But […] believing surely seems different here. I don't have voluntary control over my beliefs. Although I can simply decide to speak louder, I can't simply decide to believe or disbelieve.[15]

More specifically, the main objection to spelling out JUSTIFICATION in deontological terms is that doing so commits us to an implausible form of *doxastic voluntarism*—a thesis according to which we have voluntary control over our belief-forming processes in the specific sense that we may decide their outputs.[16] Roderick Chisholm—who, as we saw above, defends a deontological construal of justification—

[15] Kornblith (2001, p. 231).

[16] See, e.g., Alston (2005), for a recent statement of this objection.

does, indeed, frequently talk about accepting, disbelieving, suspending, and withholding judgment on propositions.[17] However, it is simply false, the objection goes, for a large and important class of beliefs—most pertinently *perceptual* beliefs—that we are able to exert such an influence on which beliefs we form.[18] Hence, since most of us are inclined to think that at least some of our perceptual beliefs are justified (or, at the very least, *can* be justified), we conclude, by way of a *reductio*, that justification does not pertain to epistemic duties—i.e., we reject the deontological construal of justification.

Let us look at this argument in more detail and start with the deontological claim from the previous section:

> (5.2.1) That *S* is epistemically justified in believing that *p* is (only) a matter of fulfilling certain epistemic duties in forming her belief that *p*.

Now, consider the following. Sitting in bed, looking to the right at my fiancé working by her computer, I simply cannot help believing that she is sitting right there, typing away. It is, in a sense, not up to me whether or not to subscribe to the belief in question—it almost seems to force itself upon me. This becomes even more obvious if we consider a more dramatic scenario where I am about to get hit by a bus: I certainly do not *decide* to believe that there is a bus coming dangerously close to me. The belief is just there, regardless of whether I want it to be or not. And even though this vivid and unyielding character certainly is no unique mark for perceptual beliefs—surely, my memory based belief that there is a street outside my window and a car in the driveway may strike me as almost as hard to voluntarily form or give up—these qualities are, undoubtedly, the most obvious in the case of perceptual beliefs.

---

[17] See, e.g., Chisholm (1977, p. 6).

[18] In the following, I will focus on the forming of beliefs and leave out the maintaining of beliefs since the reader easily can draw out the analogues herself.

In other words, it seems to be the case that,

(5.2.2) as a matter of empirical fact, we do not have the ability to voluntarily form or refrain from forming the *perceptual* beliefs we, in fact, form.

Again, it might be that the same goes for other, non-perceptual kinds of belief too but let us focus on perceptual beliefs for now and, in particular, the consequence that

(5.2.3) we are not able to choose whether or not to form beliefs in accordance with the epistemic duties that (presumably) pertain to perceptual beliefs.

Now, consider the 'ought' implies 'can' principle introduced above:

(OC) You can only have an epistemic duty to believe something if you have the ability to chose voluntarily whether or not to form a belief in accordance with that duty.[19]

The reasons to assume (OC) in the present context are best brought out by comparing it to a slightly weaker version:

[19] This kind of principle is sometimes referred to as a *principle of alternate possibilities*. If the reader is worried by Harry Frankfurt's (1969) counter-examples—launched at the classical principle of alternate possibility in ethics—in terms of circumstances that make it impossible for a person to avoid performing an action, without those circumstances in any way bringing about that he performs that action, she may use the following formulation instead: You can only have an epistemic duty to believe something if it is *not* the case that you believe that something *because* you could not have believed otherwise. Using this formulation would not alter any of the points I will be making.

(OC*) You can only have an epistemic duty to believe something if you have the ability to chose voluntarily to form a belief in accordance with that duty.

As it turns out, (OC*) is *too* weak and for the following reason: On (OC*), every belief you, *in fact*, end up with—regardless of whether you did it as a result of compulsion or the exercise of will—will be let into the dimension of deontological evaluation for the trivial reason that you always have the "ability" to believe what you, in fact, end up believing. However, it might be reasonably argued that having an epistemic duty does not merely involve having the ability to believe (or not believe) what we, *in fact*, end up believing (or not believing), but also the ability to believe *otherwise*. That is, if we are to evaluate beliefs epistemically along a deontological dimension, it must be the case that we actually *could have chosen* to believe other propositions than the ones we, in fact, ended up believing—which is captured by (OC) but not by (OC*).

To see this point clearer, consider the following example: Assume that we have a duty to never believe a falsehood. (This is clearly too demanding and unrealistic a requirement but it will still serve to prove the point.) How could one fulfill this requirement? One way to do it would be to not believe *anything*. Another way would be to only believe propositions that are proved to be true. While the former is clearly undesirable, the latter is absurdly unrealistic (and both probably even empirically impossible) but let us disregard this and instead focus on the question: Under what conditions could an epistemic agent at all be said to be *subject* to such a duty?

A not too unreasonable idea here is that an epistemic agent could only be subject to such a duty if she could, in fact, either (*a*) abstain from believing anything at all, or (*b*) chose to believe not only what she is able to prove but also to believe otherwise, even in light of such proofs. The qualification in terms of "even in light of such proofs" is important here since we do not want it to be the case that proofs would automatically lead her to believe the proved proposition (in which case she would merely satisfy the necessary condition stated

in (OC*)) but that the decision to believe what is proved is just that—a *decision*. I take it that this very idea is properly mirrored by (OC).

Returning the argument, we may infer the following from (5.2.3) and (OC):

(5.2.4) We do *not* have any duties in forming perceptual beliefs.

So, if (5.2.1) holds and JUSTIFICATION, indeed, is a deontological notion, our conclusion will have to be that

(5.2.5) our concept of justification is *inapplicable* to perceptual beliefs.

Note that JUSTIFICATION being inapplicable to perceptual beliefs does *not* imply that our perceptual beliefs are *unjustified*—only that they are *non*-justified. A belief is unjustified if it could have been justified but is not, while a belief is non-justified if the concept of justification does not apply to it in the way that the concept of righteousness or pride does not apply to a hat or a wooden box. That is, it is not just that the objects in question *happen* to not fall within the extension of the concept in this particular instance—they *could* not fall within its extension (barring metaphorical extension or substantial conceptual reconstruction).

So, what are we to make of this argument? One way to look at it is as a *reductio*, and since most epistemologists take (OC) to be a plausible principle and do not want to deny that our perceptual beliefs can be justified, (5.2.1)—i.e., the very idea that justification should be construed in deontological terms—is usually considered the odd man out. At this point, it is, of course, possible for the deontologist to simply bite the bullet, accept (5.2.5), and deny that talk about being justified applies to perceptual beliefs. However, as has been argued by Alston (among others), perceptual beliefs are in no way unique in not

being subject to any voluntary control on part of the believer.[20] Other examples that come to mind are beliefs resulting from memory and introspection. In fact, Alston goes so far as to claim that "no one ever acquires a belief at will"[21] and challenges the reader to try for herself:

> Can you, at this moment, start to believe that the Roman Empire is still in control of Western Europe, just by deciding to do so? If you find it incredible that you should be sufficiently motivated to even try to believe this, suppose that someone offers you $500 million to believe it, and that you are much more interested in the money than in believing the truth. Could you do what it takes to get that reward? Remember that we are speaking of believing *at will*. No doubt, there are things you could do that would increase the probability of your believing this [but] [c]an you switch propositional attitudes toward that proposition just by deciding to do so? It seems clear to me that I have no such power. Volitions, decision, or choosings don't hook up with propositional attitude inaugurations, just as they don't hook up with the secretion of gastric juices or with metabolism.[22]

In other words, it seems plausible to assume that an argument analogous to (5.2.1) through (5.2.5) may be constructed for other, non-perceptual beliefs, *mutatis mutandis*, with the consequence that the domain of beliefs that may be held justifiably on a deontological construal keeps shrinking to the point where *Strong Deontologism* becomes nothing short of implausible.

---

[20] See Alston (2005).

[21] Alston (2005, p. 67).

[22] Alston (2005, p. 63).

5.3. Indirect Control over Belief

At this point, the deontologist may object that the aforementioned argument presupposes a far too strong reading of "voluntary control," and that it is possible to defend deontologism about justification by recourse to a more plausible and realistic reading. More specifically, it might be objected that we have been assuming that deontologism requires *direct* control over our belief-forming processes, while it is, in fact, sufficient that we have *indirect* control or influence over the beliefs that we form.

Before evaluating the plausibility of such a claim, we need to get clearer on the taxonomy. Having *direct voluntary control* implies being able to bring about the object of control by a mere act of will, intention or volition. This is the kind of control we might have over mental imagery—i.e., our ability to bring about mental pictures—but that the previous section denied that we have over our belief-forming processes. However, there are other forms of control, most pertinently various forms of *indirect voluntary control*, located on a continuum of decreasingly direct control. For example, certain things may be brought about not by a direct act of will, but by way of a chain of events, making up a single, uninterrupted act. This is the sense in which the opening of my door or the adjustment of the temperature in my office is within my voluntary control; while I may not be able to open the door or alter the temperature merely by an act of will, the opening of the door and altering of the temperature is nevertheless within my (indirect) voluntary control, by virtue of the fact that I can bring about the necessary chain of events by walking across the room and opening the door, or turning the knob on the thermostat.

Even further out on the continuum of indirect voluntary control are the kinds of acts that may not constitute single, uninterrupted acts, like walking over and opening a door or turning a knob, but rather a series of inter-related actions spread out over an extended period of time. This is the sense in which looking for more evidence, taking steps to engage in more thorough consideration and weighing of evidence, and deliberating over how to direct my inquiry in a more strategic manner is within my (long-range) voluntary control. This is

also the kind of control Chisholm seems to be getting at in the following passage:

> If self-control is what is essential to activity, some of our beliefs, our believings, would seem to be acts. When a man deliberates and comes finally to a conclusion, his decision is as much within his control as is any other deed we attribute to him. If his conclusion was unreasonable, a conclusion he should not have accepted, we may plead with him: "But you needn't have supposed that so-and-so was true. Why didn't you take account of these other facts?" We assume that his decision is one he could have avoided and that, had he only chosen to do so, he could have made a more reasonable inference. Or, if his conclusion is not the result of a deliberate inference, we may say, "But if you had only stopped to think", implying that, had he chosen, he could have stopped to think. We suppose, as we do whenever we apply our ethical or moral predicates, that there was something else the agent could have done instead.[23]

As pointed out by Alston, however, this attempt to frame doxastic voluntarism in terms of indirect voluntary control suffers from a problem, namely that of failing to acknowledge the distinction between doing *C* to voluntarily bring about an *E*, and doing *C* to voluntarily bring about a *definite E*. Clearly, while whether or not I look out my window to see if it is sunny may be within my voluntary control, whether or not I believe what I see when I look out is not. The same goes for looking for evidence, and other inquiry related activities:

> In order that the phenomenon of looking for more evidence would show that we have voluntary control over propositional attitudes, it would have to be the case that the search for evidence was undertaken with the intention of taking up a *certain*

[23] Chisholm (1968, p. 224).

> attitude toward a *specific* proposition. For only in that case would it have any tendency to show that we have exercised voluntary control over *what* proposition attitude we come to have.[24]

In other words, the fact that we do have voluntary control over many of our actions that might give rise to beliefs, does not show that we have voluntary control over *what* beliefs we form as a result of those actions. The same goes for various forms of *indirect voluntary influence*—a kind of influence even further out on the continuum of decreasing directness—where we may take steps that either bring to bear on (*a*) candidates for belief or (*b*) our general belief-forming habits and tendencies. Indeed, as the final chapter of the present study argues, there are quite a few things we may do to alter our belief-forming habits in ways that that would improve our chances to reason successfully, i.e., to reason in a way that will tend to give us true beliefs in significant matters. However, this does *not* show that *doxastic* voluntarism is true, i.e., that we may voluntarily choose *what* to believe, as in what *specific* proposition to believe, anymore than our ability to open our eyes and look out a window on a sunny day shows that we may choose whether or not to believe that the sun is shining.

5.4. Weak Deontologism and Role Oughts

In light of this, the most natural response available to the deontologist is to deny (OC). I will now argue that material for such a move can be found in two papers by Richard Feldman, although it should be pointed out that Feldman merely sets out to defend the claim that deontological judgments are sometimes true, without committing himself to the idea that justification should be understood in deontological terms.[25] Nevertheless, a defense of what I will refer to as *Weak Deontologism*—i.e., a reading of Deontologism that relinquishes

---

[24] Alston (2005, p. 70; emphasis in original).

[25] See Feldman (2000, 2001).

(OC)—may utilize some of his arguments, and it is the purpose of this section to evaluate the prospects of doing so.

Now, the argument Feldman sets out to address is the following:

(5.4.1) If deontological judgments about beliefs are true, then people have voluntary control over their beliefs.
(5.4.2) People do not have voluntary control over their beliefs, hence
(5.4.3) Deontological judgments about beliefs are not true.

If the applicability of deontological judgments is understood in relation to (OC), the argument goes through—at least if what was argued in the previous section is correct. So, an obvious alternative strategy is to investigate whether there is an alternative reading of 'deontological judgment' such that (5.4.1) comes out false.

This is also the route taken by Feldman, who suggests that deontological judgments are to be understood in relation to "role oughts":

> There are oughts that result from one's playing a certain role or having a certain position. Teachers ought to explain things clearly. Parents ought to take care of their kids. Cyclists ought to move in various ways. Incompetent teachers, incapable parents, and untrained cyclists may be unable to do what they ought to do. Similarly, I'd say, forming beliefs is something people do. That is, we form beliefs in response to our experiences in the world. Anyone engaged in this activity ought to do it right. […] I suggest that epistemic oughts are of this sort—they describe the right way to play a certain role.[26]

For our purposes, it does not matter so much that the role of a believer and that of a cyclist, parent or teacher are different in that the

[26] Feldman (2000, p. 676).

latter are optional in a sense that the first one is not. Supposedly, while being a cyclist, parent or teacher are (largely social) roles that we may take or give up more or less as we please, there is no way we can opt out of the role as a believer. As was stressed in the previous sections, we do not choose to believe. In fact, it might even be argued that our tendency to form beliefs (i.e., our tendency to *at all* form beliefs—not what *particular* beliefs we form) is a hard-wired feature of our biological make up. However, this merely implies that the scope of the roles are different—indeed, that some roles are all-encompassing in a way that other roles are not—and is completely irrelevant as for the general implication from playing certain roles to having certain duties, which is what concerns us here.

Now, in a more recent paper, Feldman spells out what determines the right way to play a role, and how it relates not to what is normal or expected but rather to good performance:

> What counts as good performance in a role, and thus determines how a role ought to be carried out, may be dependent in certain ways on what people are generally able to do. Consider, for example, the claim that teachers ought to explain things clearly. Arguably, what counts as a clear explanation is dependent at least in part on what people are able to say and what people are able to understand. One could imagine standards for clear explanation that are so demanding that no one could ever meet them. It is not true that teachers ought to explain things that clearly. Similarly, the standards of good parenting or good cycling that apply to us are not at super-human levels. It's not true that parents or cyclists ought to do things that would require them to exceed the sorts of capacities people have. It is consistent with this, however, that an individual ought to do things that he or she is not able to do. An inarticulate teacher may simply be unable to explain things as clearly as he ought, and he may not have the capability of learning to explain things clearly. Thus, even if the standards for good performance in a role are in some way limited by the capacities

> of those who fill the role, it is not the case that the existence of those standards implies that individuals must have basic or nonbasic control over that behavior that is judged by those standards.[27]

By identifying a middle-ground for deontological judgments between the ideal of good performance and the actualities of (often highly imperfect) human capabilities, the kind of oughts that flow from inhabiting roles do not require voluntary control over the behavior evaluated. Hence, Feldman's suggestion to understand deontology in terms of role oughts, undoubtedly, carries some promise for the deontologist that wishes to hold on to her thesis by rejecting (OC). However, as I will now argue, the suggestion does not take her all the way to a viable notion of duty, at least not as far as justification is concerned.

To see this, consider the following scheme, characterizing what it is to have a duty on Feldman's account:

> *From Roles to Duties*
> If *S*, *qua* φ-er, plays the role of someone φ-ing, and φ-ers typically are able to φ in a Q manner and, thereby, meet the standard of good φ-ing, then *S* ought to (i.e., has a duty to) φ in a Q manner.

By way of example:

> If *S*, *qua* teacher, plays the role of someone teaching, and teachers typically are able to explain things in a pedagogical manner and, thereby, meet the standards of good teaching, then *S* ought to (i.e., has a duty to) explain things in a pedagogical manner.

---

[27] Feldman (2001, p. 88).

Now, as noted above, Feldman does not wish to commit himself to the deontological thesis that justification should be understood in terms of epistemic duties. Nevertheless, consider the following instance of the above schema, in terms of the biological role of believers:

> If *S*, *qua* believer, plays the role of someone believing, and believers typically are able to believe in a *Q* manner and, thereby, meet the standards of good believing, then *S* ought to (i.e., has a duty to) believe in a *Q* manner.

In order to get the kind of deontologism we are after, however, what it is to believe in a *Q* manner here has to be spelled out in terms of justification. That is, if being justified consists in fulfilling one's epistemic duties, the following would have to be the case:

> If *S*, *qua* believer, plays the role of someone believing, and believers typically are able to believe justifiably and, thereby, meet the standards of good believing, then *S* ought to (i.e., has a duty to) believe justifiably.[28]

The problem with this move, however, is that the notion of duty generated by such a schema is too weak to deserve being associated with genuine duties, in that it is too far removed from questions of praise and blame. To see this, consider the following instances of the *From Roles to Duties Scheme*, in terms of yet another biological role:[29]

> If *S*, *qua* green plant, plays the role of a photosynthesizer, and green plants typically are able to use sunlight to synthesize

---

[28] Note that this does not imply that to fulfill one's duties consists in being justified. The present formulation is perfectly compatible with justification being but one aspect of fulfilling one's epistemic duties.

[29] I owe the following examples to Jeff Sebo, who makes essentially the same point in his "Is a Real Ethics of Belief Possible?" (2006).

> foods from carbon dioxide and water in such a way as to meet the standards of good photosynthesis, then *S* ought to (i.e., has a duty to) synthesize foods from carbon dioxide and water in such a way that it meets the standards for good photosynthesis.

If this example seems unconvincing, consider another example:

> If *S*, *qua* food digester, plays the role of someone digesting food, and food-digesters typically are able to break down food by mechanical and enzymatic action into substances that can be used by the body in such a way as to meet the standards for good food digestion, then *S* ought to (i.e., has a duty to) digest food in such a way that she meets the standards for good food digestion.

What these two examples illustrate is how far removed the oughts generated by the *From Roles to Duties* scheme are from what we think about as duties, namely as something intimately connected to praise and blame. We would never think of reproaching a plant with failing to photosynthesize in a satisfactory way, or a person with not being able to digest properly (although we might blame her for having voluntarily acted so as to impair her digestive abilities, but that is quite another issue). Or better said: If we were to do such a thing, we would be speaking metaphorically, not literally.

For this reason, it seems reasonable to conclude that the oughts generated by the *From Roles to Duties* scheme are not deontological in that they do not specify duties, in any interesting sense of those words. They do, however, specify what is *good* in relation to the role occupied. In the terminology of chapter 2, the schema, thereby, specifies *general norms*. In the case of justification, this is due to the fact that, quite independently of whether or not we might have a duty to believe justifiably, it might still be the case that it is something *good* to believe justifiably, just like it might be *good* for a green plant to photosynthesize in one way rather than another, quite independently of

whether or not a plant can have a duty to photosynthesize thus. And as we have seen: In so far as such duties were to pertain to the formation of belief, the implausibility of doxastic voluntarism makes it reasonable to think that no such duties exist.

5.5. The Moral: Refinement or Reconstruction?

What is the moral of the above discussion? So far, we have been concerned with evaluating the viability of deontologism as a theory of justification, as brought out through a certain stereotype inherited from Locke and Descartes and defended in more recent years by Chisholm. This stereotype does not necessarily provide a fully accurate story about the phenomena referred to by the corresponding concept. Nevertheless, a plausible story about the origin and evolution of the stereotype may still provide substantial hints as for the kind of situation in which the phenomenon in question was baptized and—especially given what we might have found out about the world since the term was originally introduced.

We started out this chapter with one such story, involving Locke and Descartes' emphasis on how justification pertains to the fulfillment of epistemic duties—an idea that we came to dub *Strong Deontologism*. Furthermore, we saw that (*a*) subsequent research provides good reason to think that such deontologism does not provide an accurate story of its referent, in assuming that belief-formation is under our voluntary control, and that (*b*) the most promising non-voluntarist reformulation of the theory, or what we referred to as *Weak Deontologism*, does not yield a substantial enough notion of epistemic duty.

In accordance with the methodology introduced in the first part of this study, two epistemological routes are now open: On the one hand, we may *refine* JUSTIFICATION by rejecting voluntarism as a false assumption about the phenomenon referred to, while retaining the idea that justification pertains to belief-forming processes. The most promising candidate for fleshing out such a refinement would most likely be *process reliabilism*, according to which a person's degree

of justification in holding a belief to be true is a function of the reliability of the belief-forming processes that are casually responsible for the belief in question.[30] However, while reliabilism will turn out to play an important role later on in this study, we need to note that opting for such view already at this point would have to rest on an unwarranted assumption, namely that the referent inherited from our deontological predecessors does, in fact, provide the most promising means to achieve our epistemic goal of attaining and maintaining true belief in significant matter.

For this reason, the next chapter will look into an alternative route, attempting to retain voluntarism while identifying a more plausible candidate for voluntary control and, thereby, *reconstruct* JUSTIFICATION to fit better with what we now know about the involuntary character of belief-formation. The suggestion in question is far from new, however, and may be traced back to Descartes' ideas about justification as pertaining to what we may see clearly and distinctly on *introspection*. The suggestion will eventually be rejected due to an overly optimistic picture of our introspective access to the grounds for our beliefs, which, in chapter 7, will motivate an alternative reconstruction, more closely related to reliabilism than to either Deontologism or what I will call introspection-based access internalism. Still, a thorough inquiry into the suggestion as well as the reasons for its failure will provide us with valuable information about what would have to constitute a concept that serves us better.

5.6. CONCLUSION

In the present chapter it was argued that proponents of so-called *Strong Deontologism* about justification have to face up to the dilemma of either radically restricting the scope of JUSTIFICATION in face of involuntarism or giving up their position in favor of *Weak Deontologism*. The latter, however, is at the cost of leaving behind the discourse of epistemic praise and blame—a discourse that, arguably, provides

[30] See, e.g., Goldman (1986).

the reasons for sympathizing with Deontologism about justification in the first place. In the next chapter, we will look into and eventually reject an account that, in the spirit of deontologism, seeks to spell out JUSTIFICATION in terms of factors within our voluntary control but that, *contra* doxastic voluntarism, does not locate the proper object of epistemic evaluation in the holding of beliefs but rather in certain voluntary actions pertaining to the mind's awareness of itself via acts of *introspection*.

Chapter 6.
Introspection-Based Access Internalism

In the present chapter, I will consider a construal—or a reconstruction, if you will—to the effect that (*a*) cognitive accessibility should be equated with introspective accessibility, and (*b*) justification is a matter of having paid due attention to what is thus accessible via acts of introspection. I will then go on to review some results within cognitive psychology that give us reason to believe that our faculty of introspection is not a very reliable guide to the etiology of our beliefs. As such the construal not only fails to delimit an epistemically relevant notion of cognitive accessibility, but also falls short of presenting a viable concept of justification.

6.1. On Guidance and Accessibility

There is an important sense in which the Guidance Conception of Epistemology discussed above is perfectly divorceable from Deontologism. Or so I will now argue.

First, remember what GC said:

The Guidance Conception of Epistemology (GC)
The main epistemological desideratum is to *guide* epistemic inquiry, in providing means for epistemic inquirers to fulfill their desiderata.

In want of epistemic duties, and consequently any plausible desiderata pertaining to the fulfillment of such duties, what are the relevant epistemic desiderata that epistemologists are to address? Traditionally, a desideratum deemed central and important to epistemological inquiry, quite independently of the issue of epistemic duties, has been to provide means for the epistemic inquirer to answer the following question:

(Q2) Do I, *qua* epistemic agent, have any good reasons to believe that any of my beliefs about the world are true?

This is the question facing the epistemic agent deliberating about what to believe from a first-person point of view, and the challenge posed by it is, according to John Pollock, "the fundamental problem of epistemology."[1] As the reader surely notes, one *prima facie* plausible rationale for this being so is GC, i.e., the very the idea that epistemology should *guide* epistemic conduct.

More than this, many epistemologists take GC and (Q2) to be linked to a particular kind of *internalism*. I take the line of reasoning to be the following: If a central task of epistemology should be to facilitate the epistemic life of epistemic inquirers by providing means to answer (Q2)—supposedly an important desideratum for epistemic inquirers—these means better be usable for such inquirers. Hence, the following plausible, minimal constraint on analyses of epistemic justification:

[1] Pollock (1986, p. 10).

The Accessibility Constraint (AC)
Only cognitively accessible factors are relevant to degree of justification.

Consequently, we find Laurence BonJour claiming that "the justifying reason for a basic belief, or indeed for any belief, must somehow be *cognitively available* to the believer himself, within his cognitive grasp or ken," because it makes no sense that "a reason that is unavailable to [a] person be even relevant to the epistemic justification of *his* belief."[2]

However, it remains to be specified exactly *what* kinds of factors are cognitively accessible. Here, epistemological discussion has been conducted against the background of a substantial (but rarely motivated) assumption, summed up in the following thesis:

The Cognitive-Introspective Thesis (CIT)
As far as the analysis of justification goes, the epistemologically relevant sense of cognitive accessibility is that of *introspective* accessibility.

This connection between cognitive accessibility and introspective accessibility is brought out nicely in Sven Bernecker's recent definition of internalism:

> In its broadest formulation, internalism about justification (or access internalism) is the view that all of the factors required for a belief to be justified must be cognitively accessible to the subject and thus internal to his mind. Something is internal to one's mind so long as one is aware of it or could be aware of it merely by reflecting.[3]

---

[2] BonJour and Sosa (2003, pp. 24-25; first emphasis mine, second emphasis in original).

[3] Bernecker (2006, p. 81).

Hence, we find Earl Conee spelling out his particular brand of evidentialism by claiming that a person "has a justified belief only if the person has reflective access to evidence that the belief is true."[4] Similarly, Robert Audi claims that "justification is grounded entirely in what is internal to the mind, in a sense implying that it is accessible to introspection or reflection."[5] Putting the point in slightly different but recently more popular terms, John Gibbons defines internalism as a supervenience thesis, to the effect that "justification supervenes on introspectively accessible properties of the believer."[6]

A similar idea can be found in Richard Foley's theory of rationality, reconstructed by Alston as follows:

> It is epistemically rational for a person to believe that *p* at *t iff S* would believe, *on sufficiently careful reflection*, that believing *p* at *t* is an effective way of realizing the goal of now believing those propositions that are true and now not believing those propositions that are false.[7]

In Foley's own words, that analysans will be satisfied if and only if *S*

> […] has an uncontroversial argument for *p*, an argument that he would regard as likely to be truth preserving *were he to be appropriately reflective*, and an argument whose premises *he would uncover no good reasons to be suspicious of were he to be appropriately reflective*.[8]

It should be noted that Foley's analysis is slightly different from the aforementioned internalist analyses, in that it is explicitly formulated in hypothetical terms. Nevertheless, the only way to reflect on the

---

[4] Conee (1988, p. 398).

[5] Audi (1998, p. 233-4).

[6] Gibbons (2006, p. 20).

[7] Alston (2005, p. 14; my emphasis).

[8] Foley (1987, p. 66; my emphasis).

extent to which believing something is an effective way of realizing this or that goal is to first find out what you believe and, supposedly, the best way to do this is through introspection. The same goes for the identification of your arguments and their truth preserving qualities, *mutatis mutandis*. And, even though Foley does not take *actual* introspection to be a necessary condition for justification, he is still committed to the *possibility* of introspecting beliefs, arguments, and reasons.[9]

In other words, the claim is that (*a*) only cognitively accessible factors are relevant to justification (i.e., AC), (*b*) only introspectively accessible factors are cognitively accessible (empirical assumption) and, hence, (*c*) only introspectively accessible factors are relevant to justification (i.e., CIT). So, by combining AC and CIT, and assuming that only introspectively accessible factors are cognitively accessible, we get the idea that, in order to be justified, a subject must have *introspective access* to the way in which (or at least that) her reasons endow her beliefs with good reason.

6.2. Introspection-Based Access Internalism

More specifically, and against the background of these suggestions, I propose the following by way of a characterization of the stereotype pertaining to the particular kind of internalism at issue, which I will refer to as *Introspection-Based Access Internalism*:

> *Introspection-Based Access Internalism (IBAI)*
> That *S* is epistemically justified in believing that *p* means that *S*
> (*a*) has good reason for taking *p* to be true; and
> (*b*) is *introspectively aware* of the way in which (or at least aware that) her reasons endow her belief that *p* with good reason (if

[9] This might be brought out by way of an analogy: A person that never moves quickly may still be quick, as long as she has the *ability* to move quickly.

only in the Foleyan sense of a *hypothetical* awareness or reflection).[10]

As was noted by Bernecker above, IBAI is sometimes referred to as *Access Internalism*.[11] However, in order to stress the commitment not only to the access thesis AC, but also to the introspection reading of such access in terms of CIT, I have settled for the name *Introspection-Based Access Internalism*.

As for (*a*), I will understand the term "reason" as designating either a doxastic state or process (i.e., a belief or a belief-forming process) or a non-doxastic state or process (e.g., a perceptual state or process). Furthermore, I will take it to refer to an *actual* reason (epistemically neutral in the sense that it need not be a *good* reason) meaning that which *gives rise to* or *sustains* the belief in question or, differently put, that which is causally responsible for its formation or suste-

[10] As has been stressed by Alston (1985, p. 85), we need to appreciate the distinction between *being* justified in believing that *p*, in the sense of, say, certain justificatory relations obtaining between the belief that *p* and related propositions, or the belief having been produced by mental processes that are appropriate (in a, so far, unspecified sense), and *justifying* your belief that *p*, in the sense of doing something to *demonstrate* that *p*, say, actually introspecting such justificatory qualities. However, I will in the present chapter not consider the following thesis: In order to be justified, it is sufficient that certain facts about what you believe and the justificatory relations between them obtain, without there being, even in *principle*, a way of introspecting such justificatory qualities of your beliefs and their inter-relations. Indeed, I will in chapter 7 defend a version of this thesis in terms of the justificatory qualities of heuristics, but I will, *contra* Alston (1985, p. 79), not consider it a brand of *internalism about justification*, since it is clearly incompatible with the claim that all factors relevant to justification are available on introspection—a claim central to the kind of internalism I will be concerned with here.

[11] See, e.g., Alston (1986).

nance.[12] This is not to deny the possibility of *merely potential* reasons—e.g., unutilized ways of justifiably coming to believe something—but merely to narrow in on a more exact vocabulary.

In other words, the general assumption here is that reasons are a kind of cause.[13] The virtue of understanding reasons thus may be brought out as follows. First, we may note that a powerful way to explain events is with reference to their causes, because knowing the causes of events facilitates prediction and, thereby, interaction—two of the prime goals of explanation. If I know that flipping the switch will cause the light to go on, I may not only predict that, if I flip the switch, the light will go on, as well as explain the light going on with reference to that someone flipped the switch, but I can also flip the switch whenever I desire that the room be lit up.

Similarly, one important way to explain the behavior of intentional agents is with reference to their reasons, because knowing a person's reasons facilitates not only the prediction of behavior but also successful interaction. For example, if I know that people tend to approve of certain kinds of behavior because they take it to exhibit certain traits characteristic of a good character—say, honesty and bravery—I can not only predict certain situations in which behavior will be approved of and explain certain acts of approval with reference to honesty and bravery, but also see too it that I am considered as being of a good character, by acting in an honest and brave manner.

However, the only condition under which explaining the actions of intentional agents in terms of their reasons is a powerful predictive and explanatory tool is if our concept of a reason actually

---

[12] This is not to deny that a belief may have *several* reasons, in the sense that there are several factors, each of which *contributes* to the belief arising or being sustained. The points I will make could be made just as well within the framework of a more complicated picture, where a belief is held or sustained by way of a multitude of reasons. However, for simplicity's sake, I have chosen to focus on the simple case in which a belief has merely one reason.

[13] See, e.g., Dretske (1988).

tracks causes for behavior. For example, if the real cause of people's approval of certain kinds of behavior was *not* that the kind of behavior in question was (taken to be) brave and honest, but that it just happened to be the kind of behavior that always took place on a Monday, an explanation of people's dispositions to approve that was formulated in terms of honest and brave behavior would, clearly, be of little use, as far as prediction and facilitation of interaction goes. However, to the extent that our concept of a reason really *does* track causes—as in the actual causes for behavior—it will provide an invaluable cognitive tool.

This lesson carries over to epistemology, where particularly *epistemic* reasons may be construed as the doxastic or non-doxastic states or processes that give rise to (i.e., cause) and sustain belief, as was suggested above. For example, say that Paul believes that the Chrysler building is taller than 40 Wall Street, and that the cause of his belief is having been in an perceptual state that is him reading a book on New York City landmarks. What reason does Paul have for his belief? The most plausible answer would dovetail the most plausible and straightforward causal explanation: he read it in a book on New York City landmarks. In other words, the cause of his belief is the reason for his belief.

The same goes for doxastic reasons. Say that Paul also believes that the Empire State building is taller than the Chrysler building. Pondering the matter—say, if asked to list the three tallest buildings in New York City—he will, most likely, connect the inferential dots and infer that the Empire State Building is taller than 40 Wall Street. What is his reason for believing this? As above, the most plausible answer would dovetail the most plausible and straightforward causal explanation: he believes that (*a*) the Empire State building is taller than the Chrysler building, and that (*b*) the Chrysler building is taller than 40 Wall Street, as well as something to the effect that, (*c*) for any x, y, and z, if x is taller than y, and y is taller than z, then x is taller than z.

Two important qualifications are at place, however. First, it should be noted that, while all reasons are causes, not all causes are reasons. As introduced above, only the causes that give rise to or

sustain belief are epistemic reasons. More needs to be said, however. If I come to believe something as a result of, say, a chemical imbalance in my brain, then, although this is a cause of my belief, it is—arguably—not a reason for it; not even a bad reason. In other words, not just *anything* that causes a belief counts as a reason for it. And, in this sense, causality does not do the whole job. One plausible way to amend the present account is to add the following constraint: For a cause to constitute a reason, it is not sufficient that the cause is causally responsible for the formation or sustenance of the belief; it also has to be the case that the epistemic agents, if sensible, would recognize the cause as a reason, were she to become aware of it. This constraint rules out deviant causes (such as chemical imbalances), while allowing for non-deviant causes to be identified as reasons (such as perceptual mechanisms, cognitive tendencies, etc.)[14]

Note, however—and this brings me to my second qualification—that the requirement that a person has to recognize her reason as a reason, were she to become aware of it, in no way implies that a subject's reasons always are introspectively *accessible* to her. Hence, a person's actual reason may not always coincide with what she, if prompted, would *take* to be her reasons (although the above constraint would require that, if informed about her actual reasons, she would, at the very least, recognize them as reasons, granted that she is sensible). As we shall see below, it is a well-known fact within cognitive psychology that people are not always the best judges of why they hold the beliefs that they do. Nevertheless, within the bounds of the above constraint, the best *explanation* of our beliefs will consist in the citing of what actually causes us to believe what we, in fact, believe—

[14] Thanks to Helge Malmgren and Hilary Kornblith for calling my attention to this issue. It should be noted that, invoking this constraint in a *definition* of reason, would make that definition circular (since the constraint talks about what a person would recognize as a reason). However, it may still be a constraint on the concept of a reason at issue, and that is exactly how I intend to use it.

quite independently of what we *take* to be our reasons. This point will be further exemplified and elaborated on below.

Having thus clarified what it is for something to be a reason, we may turn to (*b*) and say something about how we should understand the term "introspection." In its most general sense, the term designates the mind's awareness of itself. However, I will use the term in a somewhat restricted sense to designate the mind's awareness of its own propositional attitudes (i.e., *that* something is believed), propositional content (i.e., *what* is believed), and perceptual states and processes. Furthermore, introspection is, in some sense, *inner*, in that it is not identical to any of the perceptual processes that give us information about the external world. Hence, as noted already in the above, introspection is best characterized as a *non-empirical* pathway to knowledge or justification (which, as we also saw, does not imply that it yields *a priori* warrant). Lastly, it should be noted that introspection might be but *one* pathway to objects of knowledge or justification. In other words, the present characterization of introspection is compatible with the possibility that we may come to know things through introspection that we might, just as well, have come to know through non-introspective methods.

Now, consider the following degrees of introspective transparency:

(*i*) The agent has introspective access to *what* she believes, i.e., to the propositional content of her beliefs.

(*ii*) The agent has introspective access to her own propositional attitudes, i.e., to *that* she believes this or that.[15]

(*iii*) The agent has introspective access to the *reasons* for her beliefs, i.e., to *why* she holds the beliefs that she does.[16]

---

[15] That a person has access to the (first-order) propositional *content* of her belief that *p* does not imply that she has access to the (second-order) propositional *attitude* in question, i.e., to the fact that she believes that *p*. Hence, (*i*) does not imply (*ii*).

(*iv*) The agent has introspective access to the *justificatory qualities* of her reasons, i.e., to *whether* and *to what extent* her reasons actually justify her beliefs.

IBAI assumes that we are capable of all four degrees of introspection and, furthermore, that this capability of ours is constitutive of epistemic justification. If the latter is to be the case, and as has been pointed out by Goldman, it is reasonable to assume that the introspective accessibility in question must be *reliable* and *powerful*.[17] If not, it would, clearly, be useless as a means to ascertaining whether you are fulfilling the requirements set upon you as an epistemic being. So, let us posit two requirements on the *kind* of introspective access that IBAI would be committed to, to the effect that it typically would have to be the case that,

> *Introspective power*
> when the subject reflects about the justificatory qualities of her reasons by way of introspection, a belief about the presence, absence or degree of justification is generated, and
>
> *Introspective reliability*
> when such a belief is generated, it is usually true.[18]

In the following, I will refer to the combination of introspective power and reliability as *introspective effectiveness*, the lack of such effectiveness as introspective *frailty*, and argue that if we, following Kornblith, take into account recent research in cognitive psychology, we might just have to reject the idea that we tend to have effective introspective access of type (*iii*) and (*iv*).[19] This suggests that ICT is not a

---

[16] Note that (*iii*) only collapses into (*i*) and (*ii*) under the assumption that all reasons are doxastic. As made clear above, no such assumption is made here.

[17] See Goldman (1999).

[18] See Goldman (1999, p. 275).

[19] See Kornblith (2002).

plausible way to construe cognitive accessibility and shows that IBAI, consequently, does not provide a viable way to construe justification. Or so I will now argue.

The argument will proceed in two phases. In §6.3, we will review evidence to the effect that, in many cases, our reasons for belief simply are not introspectively accessible, nor, consequently, are their epistemic merits. In §6.4, we will look into some further results from cognitive psychology suggesting that, in many cases where our reasons can, in fact, be expected to be introspectively accessible, we systematically take ourselves (and our ways of reasoning) to be passing any reasonable epistemic test with flying colors. More than this, not even in cases where we are well aware of the preponderance of such bias are we very prone to take measures to attenuate our epistemic ways, since we tend to think of ourselves as far less susceptible to bias than everyone else. I will conclude that, albeit occasionally powerful, introspection is a fairly unreliable pathway to our epistemic reasons.

6.3. Introspection and Cognitive Inaccessibility

Given that the above characterization of IBAI can be said to reveal a prevalent stereotype for justification, it is now time to delve into the actual qualities of the mechanisms that it posits and, in particular, that of a supposed effective introspective access to reasons. First off, we need to note that, according to our best psychological theories, a large part of our cognitive life takes place within the so-called adaptive unconscious and is, therefore, not within our introspective reach. So, what is the adaptive unconscious?

Actually, it is slightly misleading to talk about *the* adaptive unconscious since it is more a diverse collection of more or less independent processes and modules. Under this seemingly unitary heading, we find such different processes as those responsible for detecting patterns in our environment; attention-directing mechanisms; many instances of learning; the automatic production of feelings, preferences, and goals; the proprioceptive feedback signaling the

position of our body and limbs; as well as basic language comprehension and processes responsible for the initiation of some actions. The common characteristic of all these processes, however, is that they are all *unconscious* in the specific sense of being introspectively inaccessible. Nevertheless, they influence judgments, feelings, and behavior. Furthermore, talking about an *adaptive* unconscious is meant to convey that we are dealing with an evolutionary adaptation. Being endowed with an adaptive unconscious information processor, running parallel to a conscious system is a tremendous evolutionary advantage. We take in something in the order of 11,000,000 pieces of information per second by way of our five senses, of which we can process around 40 consciously. Hence, it goes without saying that the mind works more efficiently by delegating a great proportion of the information processing to the unconscious.[20]

Considering that a great majority of our mental life, hence, is out of our conscious reach, it is to be expected that there are many cases in which introspection yields no immediate output and simply has to give way for *inference*. In other words, if what we are trying to get at by way of introspection is simply not thus accessible, we have to infer it from other, non-introspective data. In fact, as Wilson notes, introspection (construed realistically) is in many cases a process that involves (or better said: has to be completed by) the *construction* of a narrative:

> I may not be thinking of my dentist's name right now or how I feel about root canals, but with a little introspection I can bring these thoughts and feelings to mind. No amount of introspection, however, can illuminate the content of the adaptive unconscious, no matter how hard I try. Trying to access unconscious goals and motives results not in a direct pipeline to these states, but in *a constructive process whereby the conscious self infers the nature of these states*.[21]

---

[20] See Wilson (2002).

[21] Wilson (2002, p. 163; my emphasis).

The major risk with such a constructive process is, naturally, *fabrication*. For example, it is quite common for people to overlook situational influences on their actions (for example in the form of subtle pressure and manipulation) and infer that they acted simply on the basis of their own internal states (that is, that they really wanted to do whatever they were, in fact, pressured to do)—a phenomenon dubbed the *fundamental attribution error*. Similarly, if the situational influences get too strong, people make a different but related attribution error, attributing their actions completely to the situation (say, the fact that they got paid a large sum of money to play in concert), and underestimating the extent to which they wanted to perform the behavior in the first place (e.g., that they actually really enjoy playing in concert).[22]

Such fabrication is, of course, to some extent avoidable and in many cases we are able to construct fully accurate narratives about what we want and why we act the way we do. Furthermore, a claim to the effect that there are cases in which aspects of our mental life are not accessible to us does not warrant the general claim that introspection is unreliable. More than this, it remains to be established that there is a problem of unreliable introspection in relation to particularly *epistemic* reasons, i.e., to the factors causing and sustaining belief. To warrant the latter claim, we need more empirical evidence.

However, as it turns out, contemporary cognitive psychology provides a substantial amount of evidence to this effect—evidence that recently has led Kornblith to mount a strong case against the idea that introspection provides a dependable route to the epistemic qualities of reasons.[23] One psychological phenomenon discussed by Kornblith is from a classic study by Richard Nisbett and Timothy Wilson.[24] In the study, subjects were asked to examine and rate the quality of an array of consumer goods (four nightgowns in one case, four identical

---

[22] See Wilson (2002, pp. 207-208).

[23] See Kornblith (2002). See also Weinberg (2006, p. 40).

[24] See Nisbett and Wilson (1977).

nylon pantyhose in another). In the words of Nisbett and Lee Ross, the study showed that there was

> […] a pronounced position effect on evaluations, such that the right-most garments were heavily preferred to the left-most garments. When questioned about the effect of the garments' position on their choices, virtually all subjects denied such an influence (usually with a tone of annoyance or of concern for the experimenter's sanity).[25]

In other words, the people involved in the study not only tended to (*a*) form a *belief* to the effect that the right-most garment was the best one, but also (*b*) be completely *unaware* of the influence of the relative position of the garments—i.e., of the actual reasons for their belief—and even (*c*) *deny* its influence when asked about it. And as noted by Kornblith, asking the subjects to introspect her reasons more carefully would hardly help, since few probably would have been able to at all pin-point the *actual* reason of their belief (i.e., the relative positions of the garments). If anything, introspection might even have made the (supposedly annoyed) subjects even surer that they were reasoning correctly:

> Far from helping in the process of self-correction, introspection here merely results in a more confident, though no less misguided agent […] What the experiment tells us is that he will take himself to have objectively good reasons for his belief; he will take himself to have noticed features of the nightgown on the right that make it the best of the lot […] Introspection here is powerless to detect the error made, and when called into service as a source of epistemic improvement it merely serves to certify a misguided process of belief acquisition.[26]

---

[25] Nisbett and Ross (1980, p. 207).

[26] Kornblith (2002, p. 112-113).

It is important, however, to be clear on what Nisbett and Wilson's experiment does *not* show. As pointed out by Wilson in his later book *Strangers to Ourselves*, the moral of the experiment is *not* that we never have introspective access to the true reasons behind our actions and evaluations.[27] Rather, the position effect is an instance of the more general fact that, to the extent that people's beliefs are caused by the adaptive unconscious, we might very well have access to the *results* (e.g., the belief that one pantyhose is better than another) but not always the *causes* (e.g., that one pantyhose was positioned in a particular way in relation to the others)—or at least not *qua* cases. In other words, as far as the adaptive unconsciousness goes, the claim is that we are capable of introspection of type (*i*) and probably also (*ii*), but not (*iii*) nor, consequently, (*iv*). In many cases, Wilson argues, our reasons are simply not introspectively accessible and must therefore be *inferred*. We will return to this claim in a minute.

But first we need to note two things, namely that (*a*) our reasons often pertain to the adaptive unconsciousness, and, (*b*) although the adaptive unconscious might be ever so valuable in many situations (and sometimes even necessary for our survival), it is not always to the point. In fact, studies like the one by Nisbett and Wilson indicate that it sometimes leads us astray. Another classic study discussed by Kornblith, is one by Amos Tversky and Daniel Kahneman, revealing a so-called *anchoring effect*.[28] In the study, subjects were asked for the percentage of African members of the United Nations. As the question was posed, a roulette wheel was spun and the subjects were then asked if the number that turned up was too high or too low. What Tversky and Kahneman found was that, in cases where the roulette wheel provided an anchor of 10, the mean estimate was 25, and in cases where it provided an anchor of 65, the mean estimate was 45.

In other words, the result of the roulette wheel clearly influenced the judgments of the subjects. The reason for the subjects'

---

[27] See Wilson (2002).

[28] See Tversky and Kahneman (1974).

beliefs about the percentage of African members in the UN, were, in part and unbeknownst to the subjects, constituted by the fact that the roulette wheel provided a certain anchor. According to Tversky and Kahneman, however, none of the subjects showed any awareness of the influence of the anchor provided by the roulette wheel and a fair guess is that they would simply react to a suggestion that it affected their approximation with denial or even annoyance.

These kinds of results are in no way scarce. Similar anchoring effects have since also been found in people's risk assessments, where people's estimations are thoroughly influenced by the first datum that they are provided with (say, the 50,000 annual deaths from motor vehicle accidents or the 1,000 annual deaths from electrocution),[29] price estimations, even in cases where subjects are provided with preposterously extreme anchor values ("Is the average price of a college textbook more or less than $7,128?"),[30] as well as in civil tort lawsuit situations, where the amount of money requested by the plaintiff has been shown to anchor mock juror's decisions to such an extent that the researchers studying them came to title their report "The More You Ask For, The More You Get."[31] Finally, Stacey Swain, Joshua Alexander, and Jonathan Weinberg have found evidence to the effect that people's intuitive categorization judgments in response to hypothetical examples, of the sort typically employed by philosophers in evaluating philosophical theories, vary according to whether and what other thought experiments are considered first.[32]

In sum, the common characteristic of all of these studies is the presence of factors that (*a*) either *are* irrelevant for uncontroversial reasons (the number provided by a roulette wheel, the first datum or hypothetical example presented, the preposterous anchor value pre-

---

[29] See Fischoff (2002, pp. 737-738).

[30] See Quattrone *et al.* (manuscript), discussed in Hastie and Dawes (2001, p. 103).

[31] See Chapman and Bornstein (1996). See also Hastie, Schkade, and Payne (1999).

[32] See Swain, Alexander, and Weinberg (manuscript).

sented in a question) or *should* be deemed largely irrelevant for prudential reasons (the amount of money requested by the plaintiff), but that, nevertheless, (*b*) have a clear influence on subject's judgment in a way that the subject is systematically unaware of—even on introspection.

Considering what we today know about the adaptive unconscious, and the integral although inaccessible role it plays in our cognitive life, this should come as no surprise: In many cases, the relevant factors simply are not introspectively accessible. And since they are not introspectively accessible, the subject, clearly, can not be expected to scrutinize them so as to determine their epistemic merits (or lack thereof), in the way that the IBAI theorists wants her to do.

6.4. Introspection and Cognitive Bias

Clearly, it would be preposterous to claim that we *never* have introspective access to the reasons for our beliefs—it is beyond doubt that we sometimes do. However, this does not let the IBAI theorist off the hook. In many cases where our reasons might be ever so accessible, we nevertheless fail to evaluate them correctly, not due to simple performance errors or oversight, but due to systematic cognitive bias.

To see this, let us turn to studies of hypothesis testing, of which Kornblith discusses two cases. In the first case—due to a study by Peter Wason—subjects were provided with a sequence of three numbers and told that it conformed to a general rule. Subjects were then asked to produce three sequences that, they would be told, either conformed or did not conform to this rule.[33] Almost invariably, subjects had a rule in mind that they tried, but only examined confirming instances of it. In other words, say that the sequence provided was "2, 5, 8." What Wason found was that people tended to (*a*) have a hypothesis about the general rule in mind—say, "add three to the previous number"—but (*b*) only examine confirming instances of the rule, such as "11, 14, 17", but virtually never any *falsifying* sequences, such

[33] See Wason (1960).

as "5, 8, 12" (the latter consistent with the rule "add three if the previous number is a prime; otherwise, add four"). More than this, they (*c*) tended to *hold on* to the rule they originally had in mind, *even when told that it was incorrect*. Kornblith writes:

> When a number of confirming instances were piled up, they would announce that they had found the rule governing the initial sequence. Strangely, when subjects were told that they had not discovered the rule, in more than half the cases the next sequence tested was an instance of the very rule they had just been told was incorrect.[34]

A similar kind of confirmation bias has been studied within the realm of social psychology by Lord, Ross, and Lepper.[35] In their study, two groups of subjects were recruited, one consisting of people who believed that capital punishment has deterrent effects, and one consisting of people who believed that it does not. The groups then got to read two studies, one that supported the claim that capital punishment has deterrent effects, and one that supported the opposite claim. What the researchers found was that subjects from the first group considered the first study to be well conducted and convincing but the latter highly flawed, while the second group took the second study to be well conducted and convincing but the first one to be highly flawed. In general, the studies that ran contrary to the subjects' views had a minor effect on them, while the studies that supported their views only served to strengthen them further. As noted by Nisbett and Ross, this result provides a quite interesting perspective on the influence of scientific studies on public opinion:

> Before the advent of modern social science, many questions, like the issue of the deterrent value of capital punishment, were ones for which there really was no empirical evidence

[34] Kornblith (2002, p. 117).

[35] See Lord, Ross, and Lepper (1979).

> one way or the other […] One might expect, though, that once genuine empirical evidence for such questions became available, that evidence would sway opinion to whichever side it supported or, if the evidence were mixed, that it would serve to moderate opposing views. Instead, the effect of introducing mixed evidence may be to *polarize* public opinion, with proponents of each side picking and choosing from the evidence so as to bolster their initial opinions.[36]

It should be noted, however, that this and similar "polarization biases" are in no way impossible to overcome. In a study by Puccio and Ross, it was found that a particularly efficient way to attenuate polarization was to invite people to present what they considered being the best argument for a position that was incompatible with their on stance on an issue.[37] It is a problem in this context, however, that most people tend to radically underestimate the extent to which they (as opposed to everyone else) might suffer from cognitive bias. Pronin, Lin, and Ross asked a group of Stanford students to what extent they and the "average American" displayed a series of inferential and judgmental biases, including biased assimilation of information to preexisting beliefs or preconditions.[38] The students invariably took themselves to display each of the biases to a lesser degree than the average American. To rule out that this was not just an indication of a particularly arrogant Stanford student population, the researchers conducted a similar study at the San Francisco Airport (the only difference being that they included some additional biases) only to find their earlier results replicated.[39]

Let us recapitulate. First, and as for introspective access of type (*iii*) and (*iv*), we reviewed some cognitive psychological studies indicating that processes pertaining to the adaptive unconscious are

---

[36] Nisbett and Ross (1980, p. 171).

[37] Manuscript discussed in Pronin, Puccio, and Ross (2002).

[38] See Pronin, Lin, and Ross (2002).

[39] See also Armor (1998) for similar results.

very rarely—if ever—introspectively accessible, but, nevertheless, not only influence our beliefs and judgments in substantive ways but, in many cases, are responsible for the actual reasons for our beliefs. Second, the studies on confirmation biases just discussed give us reason to believe that we, in general, are not very prone to take in evidence that runs contrary to what we believe. As Kornblith has put the point, the mechanisms involved in introspective reflection sometimes act as "sub-personal cognitive yes-men."[40] In other words, not even if we *did* have complete type (*iii*) and (*iv*) access to the processes pertaining to our grounds for beliefs, would we necessarily be particularly prone to revise them in cases where we encountered (potential) indications to the effect that our beliefs were not based on good reasons. Third, and finally, the studies on the extent to which we deem ourselves to suffer from various sorts of cognitive biases (such as confirmation bias) indicated that we, even in cases where we might be very well aware of the prevalence of cognitive bias, tend to take ourselves to be exceptions to the rule and, hence, see no reason to take precautionary measures.

When pondering the implications for IBAI, it is important to remember what we are concerned with here. In the context of, say, a researcher working with introspective data, all phenomena discussed so far may be taken into account in the evaluation of the introspective reports and dealt with through a series of coping strategies (only using phenomenologically trained subjects, getting clearer on dissociations between experience and report, weighing the introspective reports against other, non-introspective sources of data, etc.). Not so in the kind of non-lab environment that the defenders of IBAI typically wants us to introspectively scrutinize the reasons for our beliefs—an environment often void of any substantial methodological precautions. Remember that justification, according to IBAI, essentially *consists* in the (actual or hypothetical) introspection of reasons and their soundness, thereby presupposing an effective type (*iii*) and (*iv*) access to the reasons for our beliefs.

---

[40] See Kornblith (forthcoming b).

However, the surveyed experimental results warrant the conclusion that introspection does not provide a very effective route to epistemic self-evaluation and improvement, contrary to what the proponents of IBAI assume. And given what we have found out about our sometimes powerful but often quite unreliable grasp of the processes associated with reasons for belief, it seems implausible that introspective accessibility of reasons should be constitutive of something as truth-directed as epistemic justification. Quite often, we simply do not have an accurate idea of the etiology of our beliefs and, in many of the cases where we do, we would simply not identify our reasons as anything but sound.

6.5. A CARTESIAN LEGACY?

It is time to diagnose what is going on here and draw out the general implications for epistemology. Let us do this by first contrasting the picture drawn of the adaptive unconscious above with Descartes' idea of the mind. Antonio Damasio has dubbed Descartes' strict separation of the mind from the body "Descartes' error."[41] As pointed out by Wilson, however, Descartes made a related error in that he, furthermore, restricted the mental to the conscious. For Descartes, there was nothing in the mind that was not, in principle, accessible to the thinking subject and, consequently, he rejected the very idea of an unconscious—an error that Arthur Koestler has referred to as the "Cartesian catastrophe" that led to "an impoverishment of psychology that it took three centuries to remedy."[42]

Is it plausible to assume that IBAI is a mark of this very Cartesian legacy—a legacy that has remained unchallenged within large parts of modern epistemology (unlike its analogue within the theory of mind), perhaps not explicitly, but by way of its influence on the way contemporary epistemologists construe justification? I think so. For one thing, it would explain not only the emergence but also the

---

[41] See Damasio (1994).

[42] See Koestler's introduction in Whyte (1978).

significance of the model of the mind that we find within IBAI, as an arena completely transparent to the introspecting epistemic agent, not only with respect to the contents of her beliefs but also to the underlying reasons.

However, it does not explain why this view of the mind still holds such a prominent office in contemporary epistemology. Why, if the picture of the mind as transparent to the epistemic agent drawn by IBAI is so implausible, would anyone ever defend such a theory today? One answer is deontology. If justification is a matter of what we *ought* to believe in relation to our epistemic duties, and ought implies *can*, there has to be something we can, in fact, do about our epistemic conditions for justification to at all be a live issue. More specifically, there has to be something we can do in order to influence the degree to which we are justified or not and, in order to do that, the factors determining our justification (whatever they may turn out to be) have to be accessible (epistemologists have assumed) in the specific sense of *introspectively* accessible. Consequently, we not only need introspective access of type (*iii*) but also of type (*iv*). And, even if the deontological element is somewhat less explicit in modern IBAI, the dual Cartesian legacy, providing a picture of the mind as transparent to the epistemic agent and justification as a matter of what we ought to believe, could, undoubtedly, explain why some philosophers have come to embrace a view on introspection that, as it turns out, is completely at odds with modern cognitive psychological research.

Where does this leave the defender of IBAI? I take it that three routes are open. First, proponents of IBAI might simply bite the bullet and claim that, in all cases where subjects either cannot access their reasons or fail scrutinize them in an unbiased way, they are simply unjustified. The main problem with such a strategy, however, is that justification becomes a very rare occurrence, given the preponderance of bias and situations in which reasons are introspectively inaccessible. This also creates problems for the second route: to attempt to *refine* JUSTIFICATION by retaining the idea that justification pertains to introspective scrutiny and then identify conditions under which we *do* have effective introspective access to the grounds for our

beliefs. Unfortunately, the last thirty years of cognitive psychological research indicates that such conditions are quite rare. Hence, any attempt to construe justification in terms of such conditions would (again) imply that justification is a very rare occurrence—an undesirable consequence, to say the least.

This brings us to the third route: that of jettisoning CIT—i.e., the idea that cognitive accessibility should be identified with introspective accessibility—and providing a (more plausible) *reconstruction* of justification in terms of a voluntary act that (in contrast to the introspection) tends to yield and support true belief by pertaining to cognitively (although not necessarily introspectively) accessible factors, against the background not only of a realistic picture of our cognitive apparatus, but also of the particular purpose that JUSTIFICATION actually plays in typical epistemic situations. The next chapter attempts such a reconstruction.

6.6. CONCLUSION

In this chapter, we started out with the idea that only cognitively accessible factors are relevant to epistemic justification, and then considered the proposal that (*a*) cognitive accessibility should be understood in terms of introspective accessibility, and (*b*) justification should be taken to consist in having paid due attention to the justificatory qualities of one's reason for belief via introspection. We then reviewed evidence to the effect that we often lack introspective access to the actual reasons underlying our beliefs and, hence, also to the justificatory qualities of these reasons. Furthermore, related evidence indicates that, even in cases where we can be taken to have introspective access to our reasons, it seldom constitutes a reliable and powerful source of information about the actual etiology of our beliefs. In the next chapter, we will consider an alternative way to do justice to the desideratum of cognitive accessibility, more in line not only with what cognitive psychology teaches us about our mind, but also with the truth-directed discourse that we typically associate with epistemic justification.

Chapter 7.
Reconstructing Justification

We have just witnessed a second failed attempt to provide an analysis of justification that does justice to GC without being committed to any implausible assumptions about the world, be it pertaining to an implausibly strong voluntary control over our belief formation or an overly optimistic view of our introspective abilities. In this chapter, I will attempt my own reconstruction of a more apt concept, i.e., a concept that not only takes into account what we have found out about the qualities that have—for better or worse—been taken to pertain to justification, but also the specific purpose that the corresponding concept can reasonably be expected to fill, in light of the norms in which it figures and the goals that endow it with normative force.

7.1. A Minimal Notion of Justification

When attempting to diagnose the failure of IBAI, it lies close at hand to ask oneself: Maybe the mistake was to at all assume ICT, i.e., the thesis equating the cognitively accessible with the introspectively accessible? After all, barring skeptical worries, we do have cognitive

access to a whole host of things, such as many of the properties of extra-mental objects encountered through ordinary sense perception. In particular, if we sever the intimate connection between introspective and cognitive access, we see that there are many ways in which the factors relevant to justification may be accessible. More specifically, consider the following thesis:

> *The Straightforward Accessibility Thesis (SAT)*
> At least as far as the analysis of justification goes, the epistemologically relevant sense of cognitive accessibility is that of being experientially accessible to the cognizing being through introspection, perception, retention, or any other experiental pathway to matters of facts.

SAT is in line with what has recently been pointed out by Gibbons, namely that the relevant sense of accessibility might not be so much that of introspective accessibility as that of a subject *being in a position to know*.[1] More than this, SAT is more inclusive than ICT and includes introspection as one among many kinds of accessibility. After all, there is no reason to *rule out* introspection as a possible route to reasons, as long as we acknowledge—in accordance with the psychological evidence surveyed in the previous chapter—that it plays a far less prominent role in the justification of beliefs than epistemologists have traditionally assumed.

However, we need to say something more substantial and informative than this. For one thing, we need to raise the following question: Granted that introspectively accessible factors have to take the back seat, what factors are relevant to epistemic justification on this alternative construal of "accessible"? In answering this question, it serves us well to start out with the following quite pervasive (albeit vague) idea:

---

[1] See Gibbons (2006, p. 36).

A Minimal Notion of Justification

S is epistemically justified in believing that *p* to the extent that she has good reason for taking *p* to be true.

This widespread stereotype—or better said: stereotypical component—will constitute the starting point of my reconstruction and may be extracted from a wide variety of analyses, one (as we saw above) being IBAI. Consequently, even though we might not be warranted in assuming that there is enough conceptual homogeneity to speak about "our" concept of justification, it serves us well to start out with this widely held idea that justification pertains to good reason.

Two things should be noted, however. First, this notion of justification primarily applies to agents, not beliefs. In other words, although we might at times say that a *belief* is justified, this is to be understood as short-hand for that a particular *agent* is justified in holding that particular belief to be true. Second, this notion ties justification to what many people take to be the most central goal of epistemic inquiry, i.e. truth, which is a reasonable idea considering that the most plausible way to explain the normative force pertaining to justification is in relation to this very goal (or, at least, some qualified version of it). The reasons for this were discussed in chapter 2: Normative concepts are endowed with normative force by figuring in norms aimed at fulfilling certain goals. Consequently, such concepts as KNOWLEDGE and JUSTIFICATION are different from TABLE and DOG, in that the former, unlike the latter, are not only associated with a set of semantic norms pertaining to proper word use, but also with a set of particularly *epistemic* norms according to which you *should* know or *should* be justified, since knowing or being justified is (we are assuming) conducive to certain epistemic goals.

We went some length in chapter 4 to identify the specific details of these epistemic goals, concluding that they pertain to having, attaining, and maintaining true beliefs about epistemically significant matters. This, naturally, served to re-locate the question so as to ask: What are epistemically significant matters? In an effort to answer this question, it was suggested that the epistemic significance for a person

S, ultimately, is a function of that which is conducive to addressing *S*'s interests and satisfying her desires. Furthermore, in an attempt to address the possibility of quite radical variances in the latter, we followed Kitcher in assuming that most people are subject to a *healthy curiosity* and, hence, desire to know certain truths (address certain questions, solve certain problems, etc.) rather than others.[2] In other words, there is, as a matter of empirical fact, a non-negligible overlap when it comes to what people take to be significant, which has as a consequence that certain truths (questions, problems, etc.) are quite consistently deemed more significant than others. This serves as a contingent restriction not upon what people *can* but upon what people *actually do* consider epistemically significant.

Furthermore, we noted that this does *not* imply that there is no room for normative judgments on part of the epistemologist as for the weights subjects assign to their reasons for deeming something significant. On the contrary: Since such weights are directly tied to the addressing and satisfaction of the contingent interests and desires of the subject, the extent to which these interests and desires *actually* are addressed and satisfied is open to empirical investigation and, as such, provides feedback material for the acknowledgement, rectification, and improvement of the concepts and methods that are—for better or worse—employed by the subject in question.

Now, let us return to the matter at hand, i.e., that of epistemic justification, and how we may flesh out the minimal concept in relation to the concepts reviewed in the previous chapters. First, is there is any way in which the concepts reviewed in the previous chapters may be improved? Consider refinement: If it had turned out that voluntarism was, in fact, a plausible empirical hypothesis, refinement would be in place in order to sharpen the concept at hand, so as to better serve the purpose of fulfilling our epistemic duties, perhaps by invoking a more reliabilist framework of evaluation, in light of the desiderata spelled out in chapter 4. But voluntarism did not turn out to be a plausible hypothesis. Voluntarism turned out to be false and

---

[2] See Kitcher (2001).

epistemic duty (construed thus) is something that we cannot have. That is, unless it could have been demonstrated that we have a fairly effective introspective access to the epistemic qualities of our reasons, in which case JUSTIFICATION could have been easily reconstructed so as to pertain to acts of introspection rather than belief.

However, as it turns out, we often do *not* have a particularly effective introspective access to our reasons for belief. This would not necessarily have been a problem if there were, at the same time, many situations in which we *did* have such access, in which case a simple refinement would be in place, enabling our concept to better track the situations in which our introspective access is more rather than less effective. However, as we saw in the previous chapter, the best psychological evidence lends little credibility to such optimism. This brings us back to our minimal concept and calls for a further attempt at reconstruction.

More specifically, the ameliorative methodological component introduced in §4.3 above prompts two questions relevant to the spelling out of the minimal concept by way of a reconstructive improvement of JUSTIFICATION, the first one corresponding to *Evaluation* and the second one to *Improvement*:

1. What is the specific *purpose* of JUSTIFICATION in relation to the attainment and maintenance of true belief about significant matters?
2. In light of an answer to (1), as well as what we have found in the two previous chapters, what would be a more *apt* concept of justification?

I will now address these two questions in turn.

7.2. On the Purpose of Justification

As for the first question, I will take the most natural and prevalent justificatory discourse to pertain to the various social practices involved in the exchange of information. More specifically—and in

analogy with what Craig has argued in the case of KNOWLEDGE—I suggest that the purpose of JUSTIFICATION, so far as the minimal notion goes, is to flag *appropriate* sources of information.[3] More specifically, the purpose of JUSTIFICATION is to mark certain sources (and the claims that flow from them) as appropriate grounds for belief.

Given the particularly epistemic character of justificatory discourse, and what we have argued above about the goals characteristic to epistemic inquiry, the appropriateness of such sources is, to a first approximation, to be understood in terms of conduciveness to truth.[4] By marking sources as appropriately calibrated to this goal, JUSTIFICATION serves as an important tool in the planning for action in general and the formation of hypotheses in particular, and, thereby, also in the multitude of endeavors that properly fall under the heading of epistemic inquiry. Let me try to spell this out by considering how transmission of justification is usually taken to work within three specific contexts: testimonial justification, inferential justification, and non-inferential justification.

First, take the case of testimony. When consulting someone on a significant matter, we want that someone to, at the very least, have true beliefs about the matter at hand. However, since there might be ever so many informants, but only so many good testimonial sources, we look for justification as an *indication* of truth. More specifically, given that justification is understood in terms of good reason, and good reason to believe that *p* typically implies that *p* is more likely than not that *p* (if not, we would be better off guessing) the presence of (or even better: the *providing* of) justification serves to mark an appropriate source of testimonial information. This also highlights why authorities tend to play an important role in testimonial matters;

---

[3] Cf. Craig (1990, p. 11).

[4] Note that this does *not* beg the question against the internalist. Any account of justification—be it an externalist or an internalist one—takes conduciveness to truth to be an important trait of justifying reasons, grounds, or sources, at least in so far as the account in question involves a commitment to truth being an important epistemic goal.

authorities tend to be authorities by virtue of possessing good reasons within their field of expertise (and are, in general, deprived of this label if it turns out that they do not), which is exactly why reports of authorities provide (*prima facie*) reasons for belief. So, whatever else testimonial justification turns out to consist in, it is reasonable to assume that it will, at the very least, imply a likelihood of truth. And since a subset of all truths—namely, the truths that are, in one way or another, significant to us—is what we are aiming for as epistemic beings, the corresponding concept is used to flag appropriate sources of information.

Now, consider a second form of justificatory transmission, namely transmission through inference. As for deductive inference, the case is pretty straightforward: validity together with the truth of the premises guarantees a true conclusion (barring performance errors on part of the person performing the inference, of course). For this very reason, deductive inference is an excellent source of information and rightly considered one of the most worthy candidates of justification. However, considering the slightly limited scope of genuine deductive inference and, in particular, the pervasiveness of inferences that have to be made under conditions of uncertainty and incomplete information, we also value inductive and abductive inference (or inference to the best explanation, as it is sometimes called). And the extent to which we consider such inferences justified is directly proportional to their tendency (in want of a guarantee) to yield true belief. In other words, whatever else inferential justification turns out to consist in, it is reasonable to assume that it will, at the very least, imply a likelihood (or in the deductive case: a guarantee) of truth. And again: Since a subset of all truths—namely, the *significant* truths—is what we are aiming for, we, thereby, use the concept of justification to flag an appropriate source of information.

Finally, let us consider non-inferential transmission of justification, i.e., the kind of justification that, supposedly, does not rely on an inference from other propositions. More specifically, let us consider one of the main candidates for non-inferential transmission of justification, namely the justification pertaining to our perceptual beliefs.

One influential account of perceptual belief as non-inferentially justified belief is so-called dogmatism, according to which we are *prima facie* justified in our perceptual beliefs just by virtue of undergoing or having undergone the corresponding perceptual experience.[5] The justification in question is *prima facie* in the sense that it may be *overridden* by further considerations, for example if we found reason to believe that we are subject to some kind of illusion. And this carries interesting indications as to the kind of role JUSTIFICATION plays in relation to perceptual belief since it suggests that we typically assume that our perceptual faculties will not lead us astray, unless we have positive reason to believe otherwise, in which case we re-evaluate the justification transmitted by them. Hence, and as above, whatever else perceptual justification turns out to consist in, it is reasonable to assume that it, at the very least, implies a substantial likelihood of truth, since it is exactly in the cases where we find reason to believe that there is no such likelihood that we withhold the term "justified."

Hence, I conclude that the purpose of JUSTIFICATION is to flag appropriate sources of information, as in sources that tend to produce true beliefs. However, as is highlighted by Craig's related inquiry into the purpose of KNOWLEDGE, what has been said so far fails to provide an interesting distinction between the two concepts. This, however, does not show that what has been said so far is mistaken, but merely that there is more to be said. For this reason, I will now turn to the question of how we may fill in the details of this concept of justification.

If what was argued in chapter 5 is correct, Deontologism does *not* provide a viable candidate, since it is committed to the implausible idea that we have voluntary control over our belief-formation. This suggests either of two mistakes: (*a*) justification does not pertain to duties (and, hence, does not require voluntarism), or (*b*) justification pertains to a domain that we do, as a matter of fact, have voluntary control over. In chapter 6, we considered a strategy along the latter lines, where justification was tied to acts of introspection, in accor-

[5] See, e.g., Pryor (2000).

dance with IBAI. However, the suggestion turned out to be highly problematic since we seldom have such introspective access to our reasons. So, how are we to spell out the minimal notion?

7.3. JUSTIFICATION AND HEURISTICS

My suggestion is the following: There are roughly two kinds of processes that endow agents with justification in taking some things to be true rather than others, corresponding to two extremes on a continuum. The first one corresponds to a particular kind of acts, namely acts of reasoning. The second corresponds to unconscious belief-forming processes. I will take it that the workings of the former may be properly modeled as conscious heuristics or *reasoning strategies*—or "thinking about how we may better think about the world," as Michael Bishop and J. D. Trout has put it recently[6]—and the latter as *unconscious heuristics*.

Considering what has just been argued about JUSTIFICATION playing the role of flagging appropriate sources of information, and that the appropriateness in question is most plausibly understood in relation to what we strive for as epistemic beings, namely (epistemically significant) true beliefs, the following presents a natural first approximation of a framework of justificatory evaluation:

> *S* is justified in believing that *p* to the extent that *p* is produced by heuristics that tend to produce true belief.

Several qualifications are called for (four, to be exact). First, given the above distinction between unconscious heuristics and reasoning strategies (i.e., conscious heuristics), it is possible to distinguish between two aspects of justification, namely what Goldman has referred to as *primary* and *secondary justifiedness*.[7] The former corresponds to the use of unconscious heuristics that tend to produce true belief. The

---

[6] See Bishop and Trout (2005).

[7] See Goldman (1986, p. 93).

latter is a slightly more complex condition on reasoning strategies (or what Goldman calls methods), requiring that they not only (*a*) tend to produce true belief, but also (*b*) are acquired by way of other strategies or belief-forming processes that tend to produce true belief. I will follow Goldman in taking justification to require *both* primary and secondary justifiedness (to the extent that the latter applies, of course—not all justification involves the use of reasoning strategies). This is supposed to mirror the plausible assumption that it is not only desirable to form one's beliefs by way of reliable belief-forming processes and use the right reasoning strategies, but also that one is using the right strategies *for the right reasons*. Hence, a person that employs an appropriate reasoning strategy might still be unjustified, if employed on the mere basis of, say, guesswork or wishful thinking—supposedly, two highly unreliable (unconscious) heuristics. In the following, I will refer to the belief-forming processes or strategies involved in the (conscious or unconscious) selection of (other) strategies as *strategy determining heuristics*.

Second, in the taxonomy introduced in chapter 2, all norms invoking JUSTIFICATION are *generally* normative, since they all (attempt to) designate something that is conducive to one of our main epistemic goals, i.e., truth. However, only the norms invoking JUSTIFICATION on the level of reasoning strategies are endowed with *action-guiding* normativity. In other words, the distinction between general and action-guiding normativity enables us to say not only why certain unconscious heuristics are epistemically good—they are conducive to truth, which is one of our main epistemic goals—but also in what way the general form of normativity that, thereby, applies to norms involving JUSTIFICATION in terms of appropriate unconscious heuristics differs from the kind of action-guiding normativity that applies to norms that cite reasoning strategies; while both kinds of heuristics may be epistemically good, only one of them designates factors that we may adopt, reject or revise (more or less) voluntarily. This distinction will be particularly crucial in the next chapter, where we consider different attempts to improve our epistemic outlooks by developing sound reasoning strategies.

7.4. Reliability, Power, and Epistemically Relevant Worlds

This brings us to the third qualification: What does it mean that heuristics "tend to produce true beliefs"? To a first approximation, and as hinted in previous passages, this may be captured in terms of *reliability*—a notion that we already are familiar with from the discussion of introspective access in chapter 6. Here, a *perfectly* reliable heuristic is a heuristic that is *safe* in that it generates a belief in *p* only if *p* is, in fact, true, and (imperfectly) reliable to the extent that it approximates this ideal. However, while reliability, so construed, might shield us against error, it does not combat ignorance. For this reason, epistemologists have rightly stressed the importance of invoking a second epistemic quality of heuristics, namely *power*, where a powerful heuristic is one that produces a large number of true beliefs.[8]

So, considering our epistemic goal, as brought out by (D1*) and (D2*), we may explain why we should care about justification: If we understand justification in terms of heuristics that are reliable and powerful, justification enables us to attain our epistemic goals, since reliable and powerful heuristics are such that (*a*) they generate a lot of true belief and (*b*) the set of beliefs thereby generated will contain a majority of true beliefs. In analogy with our discussion of introspective access, let us refer to heuristics that are both reliable and powerful as *effective*.[9] This motivates the following reformulation:

> *S* is justified in believing that *p* to the extent that *p* is produced by effective heuristics, i.e., heuristics (*i*) generate a lot of true beliefs, and (*ii*) generate a set of belief that contains a majority of true beliefs.

---

[8] See, e.g., Goldman (1986, chapter 6) and Henderson and Horgan (2001, p. 229).

[9] I will not address the question as to exactly how to individuate beliefs, but instead assume that natural science either is or will be able to provide criteria for making such individuations.

This brings us to the fourth qualification, consisting in a specification of the set of possible worlds in which heuristics have to satisfy conditions (*i*) and (*ii*) to be properly counted as effective. For example, demanding that they are reliable in *all* possible worlds is obviously too strong a requirement, if we also want them to be at all powerful, since only more or less non-generating processes are reliable in all possible worlds (i.e., under all circumstances). Furthermore, a lot of processes that are intuitively reliable, in that they serve our needs perfectly well in the actual world, would come out unreliable on this reading only because they would be completely useless in distant possible worlds. On the other hand, demanding that the processes only need to be reliable in the actual world would be too weak a requirement, since it would not rule out a lot of processes that are reasonably categorized as unreliable in that they would yield falsities if the world was only slightly less cooperative.

For this reason, it is desirable to delimit the set of possible worlds that are relevant to the kind of epistemic evaluation at issue. More specifically, we want to identify a set of possible worlds that is richer than the one containing only the actual world, but still leaves out worlds that are simply too bizarre to enter into our epistemic evaluations. David Henderson and Terence Horgan have suggested that what we are after here are the *epistemically relevant* possible worlds. Henderson and Horgan characterizes these worlds as those that satisfy the dual criteria of being such that the epistemic agents therein (*a*) have appearances of roughly the same character as we do in our (actual world) everyday experiences, but (*b*) are not just brains in vats governed by evil demons or malevolent scientists who make sure that the epistemic agents only receive deceptive inputs.[10]

I believe that Henderson and Horgan are on to something here but that their particular characterization is somewhat off the mark and in need of modification. As for condition (*a*), the best way to understand why we would want to restrict the set of epistemically

[10] See Henderson and Horgan (2001, p. 237).

relevant worlds to worlds in which agents have appearances of a character similar to ours is to consider the kind of situations in which we typically evaluate epistemic inquirers. Not surprisingly, those situations are, from the standpoint of what is metaphysically possible (not to mention logically possible) quite *ordinary*. For one thing, they are situations in which a certain set of natural laws hold, namely *our* natural laws. These natural laws guarantee a certain degree of stability to our epistemic conditions—a stability that enables us to interact successfully with the world by inductive means. More specifically, given a set of natural laws (and note that these laws need not be deterministic), and a set of initial conditions, some phenomena occur with a substantially greater likelihood than others. As a consequence, even given the realm of all nomologically possible worlds, we tend to find ourselves in the *subset* of the nomologically possible worlds that is made up of the worlds that are either determined or more likely to become actualized (depending on whether determinism holds or not), given the state of what, at the moment, happens to be the actual world.

In other words, there are two factors relevant to the stability that enables us to interact successfully with the world by way of inductive reasoning, pertaining to, first, a constant set of natural laws and, second, the future events (if determinism is true) or chance distributions (if determinism is false) that those laws give rise to when combined with the initial conditions provided by the state of the (actual) world. It is exactly under these conditions that we negotiate the world as epistemic inquirers, which should be reflected in our epistemic concepts. Consequently, it would not only be pointless but probably also contra-productive and costly to devise our epistemic concepts and norms for anything but these very conditions. Hence, the epistemically relevant worlds are, to a first approximation, a subset of the nomologically possible worlds, namely the subset consisting in the worlds where the initial conditions are similar to the ones in the actual world. Characterized thus, it is, furthermore, reasonable to assume that the appearances of agents in epistemically relevant worlds

will be of, roughly, the same character as ours, just like Henderson and Horgan notes.

However, nothing said so far rules out the scenarios mentioned in (*b*). To understand why someone would consider that a problem we need to get clearer on the way in which such scenarios typically are introduced. In short, they are usually introduced within the framework of a methodology not too different from DCA and under the assumption that the proper role of epistemology is to exhaustively elucidate concepts. More specifically, consider the following scenario: Alice lives in our world. As far as epistemic conduct goes, she is neither better nor worse than most of her fellow inquirers. Furthermore, her beliefs are formed in the way beliefs usually are formed here, i.e., through sense perception, reasoning on various levels of abstraction, etc. As it happens, these ways of forming beliefs are reliable in our world, in the sense that they tend to generate an epistemic track record with a larger proportion of true than of false beliefs. Beth, on the other hand, lives in another possible world. She is just as epistemically well behaved as Alice and her beliefs are formed through the same kind of cognitive faculties as Alice's. In fact, they can even be considered to be epistemic twins, having qualitatively identical cognitive input at all times. Beth, however, happens to be subject to a scientific experiment by a group of evil demons, where her brain has been hooked up to a computer that feeds her with an interactive and perfectly consistent set of appearances in response to which she comes to form beliefs that are inadvertently false about the external world. Still, many philosophers feel inclined to say that Alice and Beth, nevertheless, can be justified to the same extent, and that justification, hence, cannot be a question of mere reliability.[11]

Arguments like these only go so far, however, and, more importantly for our purposes, they do not go very far at all in this context. To see this, consider the conditions under which arguments like the one above would carry some force, namely if we (*a*) were con-

[11] See Lehrer and Cohen (1983, pp. 192-193) and BonJour and Sosa (2003, p. 27) for arguments along these lines.

cerned with elucidating a pre-existing concepts, (*b*) took an armchair probing of our categorization intuitions to be a plausible way to do so, and, furthermore, (*c*) were working under the assumption that *Exhaustiveness* was a plausible desideratum to invoke in such elucidations. As it happens, we are doing none of this here. For one thing, we are not trying to elucidate a pre-existing concept; we are trying to devise a new one that will serve us better than the ones discussed in chapters 5 and 6. For another, and for reasons discussed in chapters 1 and 2, we would not necessarily take armchair probing of categorization intuitions to be a very promising way to elucidate such pre-existing concepts if that was at all our mission, considering the far more rigorous methods of contemporary cognitive science. Finally, and as we have seen in chapter 3 and 4, there is no obvious reason why we, as far as the elucidation of concepts go, would need anything but a fairly modest and far from exhaustive accounts of our epistemic concepts—as yielded by Meaning Analysis, not traditional conceptual analysis—in order to make epistemological progress. That is why the intuitive judgments that one might or might not be prone to make in response to scenarios of the above kind are largely irrelevant to the issue at hand, and why we may simply conclude that the set of epistemically relevant worlds is the subset of the nomologically possible worlds with initial conditions similar to the ones in the actual world.

At the same time, it will serve us well to look closer at some of the scenarios that are typically taken to be problematic to the kind of account that is being defended here, if only to get a better idea of the implications of the account at issue. Before doing this, however, we need to fully spell out our reconstructed concept. More specifically, in light of what has been argued above, I take it that we may reformulate our characterization along the following lines:

> *A Reconstructed Concept of Justification (RC)*
> *S* is justified in believing that *p* to the extent that *p* is produced by effective heuristics, i.e., heuristics that, in all or a very wide sub-set of the *epistemically relevant worlds,* (*i*) generate a lot of true

beliefs, and (*ii*) generate a set of belief that will contain a majority of true beliefs.

It should be noted that RC, unlike IBAI, constitutes the outline of a kind of account that actually meets the two important desiderata that have been occupying us over the course of the present inquiry, and that provided the very standards that served to disqualify IBAI. More specifically, RC not only (*a*) pertains to *cognitively accessible* factors, such as belief outputs, doxastic track-records of belief-forming processes and reasoning strategies, etc., but also (*b*) is intimately connected to what is typically considered to be one of the main epistemic goals, namely truth.

7.5. Applying the Reconstructed Concept

Next, let us get a somewhat firmer grip on RC by considering a couple of applications, especially in relation to the supposedly problematic scenarios alluded to in the above.

First, testimonial warrant. On RC, a person, *S*, is justified in believing that *p* as a result of learning that *p* from *U* if and only if (*a*) consulting *U* on issues of the type to which *p* belongs constitutes an effective reasoning strategy, and (*b*) the unconscious heuristics involved in the formation of the belief that *p* are efficient, especially in so far as they pertain to the strategy determining heuristics at work, i.e., the heuristics responsible for *S* opting for consulting *U* in the first place. For example, say that I want to know whether the Red Sox beat the Yankees last weekend. I consult one of my colleagues and learn that, yes, the Red Sox did actually beat the Yankees. Is my belief justified? Yes, to the extent that there is nothing awry with my unconscious belief-forming heuristics, consulting my colleague on baseball related issues constitutes an effective reasoning strategy (say, my colleague would never miss a Red Sox or Yankees game and is, hence, a reliable informant when it comes to the outcome of Red Sox or Yankees games), and I chose to consult her as the result of an effective heuristic (say, as a result of having had many discussions with her

about baseball and always finding her to provide me with accurate information, and, hence, coming to trust her judgment, as opposed to simply choosing a colleague at random).

Next, let us consider inferential warrant. On RC, a person, *S*, is justified in believing that *p* as a result of inferring it from a set of other propositions by way of certain inference rule if and only if (*a*) applying the rule in question constitutes an effective reasoning strategy in relation to problems of the type to which *p* belongs, and (*b*) the unconscious heuristics involved in the formation of the belief that *p* are effective, especially in so far as they pertain to strategy determining heuristics, i.e., the heuristics involved in *S*'s choice to opt for the rule in the first place. For example, say that I am working as a sales clerk and a cash register malfunction forces me to add the prices of my costumers' items with the help of pen, paper, and long addition. Are the beliefs that I reach about the totals of my customers' items justified? Yes, to the extent that there is nothing (substantially) wrong with my unconscious belief-forming heuristics, using long addition constitutes an effective reasoning strategy to add (the relevant kind of) numbers, and I opted for long addition as a result of an effective heuristic, i.e., not by, say, simply guessing that that was an appropriate strategy to use.

Finally, let us turn to perceptual warrant. A person is justified in believing that *p* as a result of perception if and only if the belief-forming processes operative in the production of her belief that *p* constitute effective heuristics within the relevant situation. Say, for example, that I see a bird in the tree outside my office window and I, as a result, come to form a belief that there is a bird on the tree outside my office window. Am I justified in my belief? Yes, to the extent that forming perceptual beliefs about middle-sized objects under the relevant circumstances constitutes an effective heuristic. As the reader notes, there is no mentioning of any kind of reasoning strategy in the case of perception. One way to bring out the plausibility of describing perceptual warrant thus is by considering the case of the reliable clairvoyant Norman, who, unbeknownst to him, is capable of effectively coming to believe things within a specific domain—say, pertaining to

the whereabouts of the American President—by way of clairvoyance. More than this, Norman has no information whatsoever about the reliability or power of his of clairvoyance; his clairvoyant beliefs are simply spontaneously formed and seem to Norman to come from nowhere. This, some epistemologists feel inclined to say, presents a problem for any account of justification essentially defined in terms of reliability and power, since such accounts seem to commit us to saying that, due to the effectiveness of Norman's clairvoyant faculty, the resulting beliefs are justified.[12]

Before evaluating this claim, however, the example needs to be described more in depth. More specifically, it needs to be specified which of the following is the case:

(7.5.1) Norman comes to believe certain things effectively by way of his clairvoyant faculty, and does, in doing so, not rely on any reasoning strategy.

(7.5.2) Norman comes to believe certain things effectively by way of his clairvoyant faculty, and does, in doing so, rely, in part, on a reasoning strategy along the following lines: "Affirm that which comes to you spontaneously and seemingly from nowhere."

If (7.5.1) provides the correct reading of the scenario, it should be noted that we often—indeed, constantly—find ourselves in Norman-like scenarios, namely when it comes to our ordinary perceptual beliefs. After all, we most often do not know and in many cases do not even have any beliefs to the effect that our perceptual faculties are effective. Does this make all our perceptual beliefs unjustified? That seems unreasonable. After all, why would everyone short of neurological or cognitive scientific expertise, or anyone living at a time prior

[12] See BonJour (1985) for the original formulation of this case. Bonjour formulates the case exclusively in terms of reliability. If anything, my reformulation in terms of effectiveness creates an even stronger case against the reliabilist.

to anything even close to a correct, scientific description of our perceptual apparatus, be unable to form justified perceptual beliefs? I see no reason why.

If (7.5.2) provides the correct reading, however, we have to remember that RC not only requires effective belief-forming processes, but also effective reasoning strategies. Clearly, affirming that which comes to one spontaneously and seemingly from nowhere is not, in general, an effective reasoning strategy—which is exactly the kind of strategy that Norman, on this description, is following. Hence, in this case, RC yields the verdict that Norman is, in fact, unjustified, which, according to the proponents of the problem of reliable clairvoyance is the correct answer.

But what about if Norman's reliable clairvoyance was, in fact, *frail*? In that case, RC would yield the verdict that he would be unjustified. But given the above analogy between Norman and ordinary perceivers, this raises a troubling question: What if our perceptual faculties are frail? This brings us to yet another scenario that has been considered a problem for accounts of justification defined in terms of reliability and power, and that was briefly attended to above: the evil demon scenario. Remember, the two main characters in the evil demon scenario are Alice and Beth, both of whom are just as responsible, but only the former of which forms her belief by way of effective heuristics. Clearly, RC yields the verdict that Alice and Beth are *not* justified to the same degree—Alice is justified while Beth is not. I do not wish to deny that this runs contrary to some epistemologists' intuitions. However, I do wish to deny that this is a very wise intuition to hold dear. More specifically, I will do two things: First, I will identify the general assumption that yields this particular judgment in the evil demon scenario, and then, by applying these assumptions in other cases, show why they are misguided.

As for the first step, I take it that the judgment that Beth is justified stems from two assumptions: (*a*) that Anna and Beth are just as epistemically responsible, and that (*b*) that there cannot be a difference in justificatory status without a difference in epistemic responsibility. Before evaluating these assumptions, we need to better under-

stand the notion of epistemic responsibility at work here. To my mind, there are two general ways in which such responsibility can be cashed out:

> *Subjective Responsibility* (SR)
> A subject, *S*, is epistemically responsible iff she does her best to make sure that her beliefs are formed in accordance with what she *takes* to be the ideal way of forming beliefs.

> *Objective Responsibility* (OR)
> A subject, *S*, is epistemically responsible iff she does her best to make sure that her beliefs are formed in accordance with what is, in fact, the ideal way of forming beliefs (whether or not she *takes* this to constitute the ideal way of forming beliefs).

Now, assume that we are working with SR, and consider the following scenario:

> *The Diligent Guesser*
> Celia forms a large set of her beliefs by way of guessing. More than this, she diligently makes sure that everything she believes by other means, i.e., through sense perception, reasoning on various levels of abstraction, etc., is in accordance with what she takes to be the ideal way of forming beliefs, namely through guesswork. However, guesswork is an unreliable way of forming beliefs.

Is Celia justified? If we stick to SR, and hold on to (*a*) and (*b*) above, we would have to say that Celia is, in fact, justified—perhaps even more justified than Alice and Beth, namely if Celia's diligence and sense of epistemic responsibility exceeds that of Alice and Beth's. After all, she is doing her very best to make sure that her beliefs are formed in accordance with what *she* takes to be the ideal way of forming beliefs. That she happens to be gravely mistaken in this is simply

beside the point on an SR reading of responsibility. I take this to provide us with sufficient reason to reject SR as an implausible account of epistemic responsibility.

This leaves us with OR. However, on OR, Celia comes out unjustified, at least if we let the ideal way of forming beliefs be linked to that which is conducive to the formation of true belief—something that, as we have seen, lies at the very heart of epistemic evaluation. Furthermore, on OR, our evil demon victim Beth turns out to be unjustified too. In other words, on pain of subscribing to an implausible account of epistemic responsibility, namely the one that not only seems to be at work in leading some of us to believe that Beth is, in fact, justified, and but also commits us to saying that Celia is justified too, we will have to agree that degree of justification is intimately connected to the effectiveness of the underlying belief-forming processes and reasoning strategies. And this, as we have seen, is exactly what RC suggests.

7.6. How Science Solves The Generality Problem

I now turn to what some epistemologists take to be a wide-ranging problem for an account like RC, due to the fact that it is formulated in terms of *types* of heuristics. Some epistemologists—perhaps most persistently Richard Feldman and Earl Conee—have viewed this as a potential problem: How, they ask, are we to identify the particular type relevant to the evaluation of a process or method's tendency to generate true belief, given that every process or method token, like any particular anything, is an instance of indefinitely many types?[13] This is the so-called *generality problem*.

In a recent paper, Klemens Kappel makes a helpful distinction between *recognitional capacities* and *theories of determination*, where the former make up (conscious or tacit) capacities to pick out relevant types, and the latter constitute substantial theories about *why* certain

[13] See, e.g., Feldman and Conee (1998), and Feldman (1993; 1985). See also Plantinga (1988).

types are relevant, while others are not.[14] What those moved by the generality problem want to draw our attention to is that we seldom—if ever—are in possession of anything like a full-fledged theory of determination. Moreover, they take this to present a problem for any account understanding knowledge or justification (in part or in whole) in terms of the reliability or power of belief-forming processes and reasoning strategies since

(7.6.1) any theory defining justification in terms of process or method types needs some way of distinguishing the relevant from the irrelevant types, and

(7.6.2) the only way to do so is by way of a full-fledged theory of determination.

For simplicity's sake, let us focus on accounts formulated in terms of reliability, since any point made against such an account may be easily reformulated in terms of power (as will any reply to such a point). Moreover, let us refer to any such account as reliabilist. Now, why do reliabilist accounts need a way to distinguish relevant from irrelevant types? The reason is that, in want of a way to distinguish processes and methods thus, the reliabilist account runs the risk of being empty. After all, on reliabilism, a belief is, to a first approximation, justified in so far as it is generated by reliable heuristics. This, however, tells us nothing about *what* beliefs are justified, unless combined with a story about what heuristic type a particular belief-forming process or method token belongs to—i.e., what is, in fact, the *relevant* type. Hence, Feldman:

> We have no idea what the theory implies about the epistemic status of beliefs until we know which types are relevant to

[14] See Kappel (2006). Kappel actually refers to recognitional capacities as *criteria of relevance*, but I have settled for the former terminology since I consider it not only more in line with entrenched philosophical terminology, but also a more fitting term for the phenomenon at issue.

> their evaluation. Without specifying relevant types, the theory is seriously incomplete.[15]

This is the first of two aspects of the problem of generality. I will refer to this aspect as *the problem of vacuity*. This brings us to (7.6.2) and the idea that the only way to provide such a story about the relevancy of types is by way of a full-fledged theory of determination, i.e., a substantial theory about the features that make a process or method type relevant. I will eventually reject this assumption and claim that we can solve not only the problem of vacuity but also the second aspect of the generality problem—an aspect that I will refer to as *the problem of testability*—with mere recourse to certain recognitional capacities that, furthermore, seem to be at work in a particularly refined and determinate form in the natural sciences. First, however, I will spell out this second aspect by looking into and criticizing what has come to be a particularly popular way of trying to solve the generality problem, while subscribing to both (7.6.1) and (7.6.2) above, namely psychological realism.

Psychological realism tries to solve the generality problem by rejecting what seems to be an underlying assumption among many of the latter's proponents, namely that there are no objective, psychological facts of the matter that determine which type is the relevant one, i.e., the type of which the particular process or method should be deemed a token. Elaborating on Alvin Goldman's idea of cognitive processes as *functional operations* or procedures that map inputs onto outputs,[16] William Alston writes:

> […] every belief formation involves the activation of a psychologically realized *function*. That activation yields a belief with a propositional content that is a certain function of the proximate input. This function will determine both what features of the input have a bearing on the belief output and what bearing

[15] Feldman (1993, p. 41).

[16] See Goldman (1979).

> they have, that is, how the content of the belief is determined by those features.[17]

In other words, the causally operative function determines the epistemologically relevant type. The underlying assumption here is a form of *psychological realism*, an idea to the effect "that there is always a unique correct answer to the question 'What mechanism, embodying what function, was operative in the generation of this belief?'"[18] This assumption might, of course, be challenged, but it is important to note that, in order to come to terms with the generality problem, the assumption need not necessarily be that bold. For one thing, it is both compatible with a fallible access to the facts of the matter, as well as some degree of indeterminacy of psychological functions—as long as the degree of determinacy is sufficient to fix the relevant type, of course.

However, there is another problem with this solution: It runs the risk of simply relocating the problem of vacuity. To see this, note what kind of work psychological realism is supposed to do here: it is supposed to makes sure that there is always a unique and correct answer to the question of what functional operation is at work in the formation of a token belief. However, it is somewhat unclear why the generality problem cannot just as well be raised for (types of) functional operations as for belief-forming process or method types. Remember, it is the former that is supposed to determine the relevant features of the latter, and if the latter can instantiate an indefinite number of kinds, why cannot the former do the same? Functional operations are specified by proximate input and belief output, and both of these factors can be specified along the lines of a rich multitude of aspects, which is exactly the kind of conditions that got the generality problem off the ground on the level of belief-forming processes and methods. And if the mere purpose of psychological realism is to ban these kinds of question in the case of functional

[17] Alston (2005, p. 126).

[18] Alston (2005, p. 139).

operations, then why not simply cut the middleman and postulate that there is always a uniquely correct answer already at the level of belief-forming processes and methods? I cannot see why that strategy would be any less *ad hoc*.

Now, I do not want to claim that this problem is insurmountable. In fact, I believe that promising attempts have been made at solving it.[19] However, what I *do* want to claim is that we do not need psychological realism in order to deal with the generality problem, for the simple reason that we do not need to provide a theory of determination in order to come to terms with the latter—or so I will argue. More specifically, I will, in the following, deny (7.6.2) and argue that (7.6.1) may be handled without reference to anything like a theory of determination. In doing this, I will also address a potentially serious methodological problem raised by Feldman, introduced in the following paragraph:

> If every belief is an instance of many types, some reliable and some not, by picking the right type for any particular case, you can make it look like a positive instance of reliabilism or a counterexample to the theory. It may be that the theory looks good to its advocates simply because they've gerrymandered the types to make it have the right results. Similarly, reliabilism may look bad to its critics partly because they gerrymander types in ways that appear to pin incorrect implications on the theory.[20]

This is the aforementioned problem of *testability*; without a robust way of adjudicating claims about the relevancy of types, philosophers are free to define relevance in a way that suits their particular theory. However, I will deny Feldman's assumption that any strategy aiming to adjudicate such disputes on a case-by-case basis, rather than with

---

[19] See, e.g., Beebe (2004), especially what he refers to as the second part of his solution.

[20] Feldman (1993, p. 42).

reference to a theory of determination, necessarily has to be "arbitrary and ad hoc" and, hence, philosophically unacceptable.[21] Next, I will discuss the typing of ordinary, middle-sized objects and, thereby, provide a framework for an elaboration on how a more refined and consistent case-by-case adjudication, of exactly the kind that may solve generality problems, regularly and consistently takes place within the sciences in relation to the typing of mental and methodological tokens.

We may start out by noting that successful inductive inference requires not only (*a*) an ability to pick up on *regularities* but also (*b*) on-point dispositions to posit certain underlying mechanisms for some of these regularities. This is because regularities *as such* are cheap and abundant, while the ones that matter—i.e., the ones that depend on actual, non-accidental structures and relations in the world—are far more rare but also immensely more valuable to inductive creatures. So, given the obvious advantage of being able to weed out the irrelevant and accidental from this abundance, it should come as no surprise that our disposition to posit underlying mechanisms seem to be heavily constrained by what might just be innate, conceptual structures that, furthermore, are likely to mirror extra-mental structures in nature.[22]

As a consequence, we are pretty good *token typers*. That is, we are pretty good at putting things (tokens) in inductively valuable categories (types). This is certainly not to say that we enter the world with fixed and ready beliefs about the structure of the world. Quite the contrary; within this possibly innate framework, our dispositions to posit underlying mechanisms are most likely highly sensitive to new, experiental data that may serve to provide further and more fine-grained restrictions on our token typing tendencies. More specifically, our typing dispositions are most likely experientially *adaptive* through conceptual development.[23] Although this, in no way, implies that our

---

[21] Feldman (1985, p. 159).

[22] See Kornblith (1993).

[23] See, e.g., Murphy (2002) and Murphy and Medin (1985).

dispositions are introspectively *accessible*, they can, nevertheless, be expected to guide and influence our conduct in a variety of respects.

Against the background of this, scientific inquiry presents a particularly interesting interface between hard-wired adaptations and acquired heuristics. In particular, I will now argue that, within the natural sciences, our token typing abilities extend beyond the picking out of middle-sized objects and phenomena in the world, to the typing of our own cognitive means and tools. More than this, successful typing of this sort is exactly what is required to solve the generality problem, and natural science can, as such, be expected to solve generality problems on a daily basis—or so I will now argue.

Let us start by considering the typing of *methodological* tokens. For example, consider Dr. Robert Koch, who discovered the bacillus of tuberculosis. Now, let *H* be the hypothesis that tuberculosis is caused by a bacillus. In testing *H*, Koch not only relied on already existing methods—such as using bacterial growth media—but also improved upon those methods significantly. In other words, one of the first things Koch must have done is (*a*) *identified* what methods researchers had been using before him (e.g., as they pertained to the use of bacterial growth), (*b*) *evaluated* if those methods led to correct predictions and experimental success, and to what extent the domains of success may be related to the domain of *H*, and (*c*) *applied* or *revised* them correctly in relation to *H*. In doing this, Koch must have typed the token methods and, in doing so, (consciously or unconsciously) made substantial assumptions about which properties of the individual tokens should or should not count towards the type in question. The generality problem looms. So, let us ask ourselves: What kinds of factors is he likely to have included and excluded, respectively?

Given substantial background knowledge as well as certain general (and most likely implicit) assumptions about the world, he most likely included factors pertaining to experimental apparatus used (e.g., the agar plates used to culture microorganisms), theoretical assumptions made (e.g., about the conditions under which a bacillus can

at all be said to be the *cause* of a disease), and any potentially interfering factors (e.g., failure to sterilize the agar plates, or failure to prepare the nutrients, salts, and amino acids for the sterile dishes properly) while not taking into account the color of the agar plate, that the name of his wife (who suggested the use of agar plates) started with an "E," nor that the final digit of the year of the experiment in question was an even number. And even though there might not be any principled way of ruling out the latter factors as part of the type in question, and that Koch might have been ever so unable to construct a full-fledged theory of determination—again, the workings of token typing abilities might be ever so introspectively inaccessible—his background knowledge and implicit assumptions about the world still must have radically constrained what he, in the end, considered to be the relevant type.

After all, although most often excellent at science, scientists are notoriously bad at describing *how* they do science. This is to be expected since an ability to ϕ far from always implies knowledge as to *how* one ϕ-s—particularly when ϕ-ing, like doing science, pertains to a large extent to practical skills and know-how. Having said that, is Koch likely to have identified a *uniquely* correct type? Probably not and, to this extent, the generality problem might actually be raising a real, practical issue. However, it should be noted that inductive success does not require picking out *uniquely* correct types. Realistically speaking, it is likely that there will be a set of correct types that, although differing in a variety of subtle details, nevertheless, all serve the practical purposes of correct prediction and experimental success well. And if so, the interesting question is not so much whether Koch identified a uniquely correct type, as whether the following holds: Was the type identified by Koch a member of the set of correct types that are all such that they may serve the practical purposes of correct prediction and experimental success to a satisfactory degree? Given the success of his research, the answer would have to be 'yes.'

Let us now turn to the ability to type more basic, *mental* tokens. After all, in evaluating *H*, Koch must not only have typed a set of methodological tokens, but also been sensitive to a whole host of mental tokens, and, thereby, to the conditions under which he could trust his own perceptual apparatus. More specifically, in typing the mental thus, he most likely must have been sensitive to the lighting conditions in the room of the experiment as well as to any interfering factors pertaining to, say, fatigue and distractions. However, experimental success in no way required that he was sensitive to whether he happened to wear a red sweater at the time of the experiment, that his first name started with an "R," nor that the experiment took place on, say, a Thursday. In other words, even though there might not be any principled way to rule out the latter factors as part of the type in question—again, regularities *as such* are cheap—nor possible to construct a full-fledged theory of determination for the relevant mental types, his background knowledge and previous experience must have radically constrained what he, in the end, would have considered relevant, by virtue of his sensitivity to the types that served the practical purposes of accurate prediction and experimental success to a satisfactory degree.

Similar examples are abundant in the history of science. Take Tycho Brahe, for example. Brahe was one of the first astronomers to conduct exact, astronomical observations and his data were inherited by Johannes Kepler who, as a result, was able to finish the project, started by Copernicus, of providing a heliocentric theory of the solar system. Brahe famously rejected the idea that stars and other astronomical objects were attached to layered crystalline spheres, and, in conducting his observations, more or less had to revise all existing astronomical tables. Needless to say, this cannot have been a small task. When claiming that everyone else is wrong, you better check your digits and the way in which you have produced them. As such, Brahe's project must have required a minute attention to methodology, in the identification, evaluation, application, and revision of pre-

viously accepted methods (e.g., as those methods pertained to the use of astronomical tables). Furthermore, Brahe must not only have been sensitive to the (relevant) conditions under which he could trust his own perceptual apparatus (e.g., in astronomical observation), or perform the required reasoning task (e.g., in calculation and inference); he also employed state of the art equipment the proper use of which required detailed technological knowledge, hosts of experimental assumptions, as well an impressive sensitivity to possibly interfering factors. And the fact that his results not only were some fifty times more accurate than that of Jonathan Muller's—the pupil of George Purbach, the founder of observational astronomy—but also were applied with astounding success by astronomers as well as farmers, sailors, and watchmakers, provided as good reasons as any that he got many things right.[24] Needless to say, this would, most likely, have been impossible if he was unable to pick out relevant methodological and mental types.

Consider also Charles Darwin—the father of the theory of evolution and natural selection. Just like Brahe, Darwin came to construct theories that conflicted with what, in many ways, was the received view at the time. Substantial parts of his impressive set of data were collected over the course of a six-year trip over the world on *the Beagle*, but the formulation of the theory that Darwin eventually was to endorse would have been impossible if it was not also for his ability for theoretical synthesis. Almost fifty years before Darwin set about his journey on *the Beagle*, geologist James Hutton reached the conclusion that the world was, most likely, hundreds of thousands of years old—not just over six thousand, like it says in the Bible. Charles Lyell was soon to provide an, at the time, even more stunning hypothesis, when he claimed that the world might even be *millions* of years old.

This provided a new framework for explaining natural phenomena and, in particular, the possibility for small, incremental

[24] Mason (1962, p. 134).

change making a big difference in the long run. Darwin combined this framework with a principle that he essentially borrowed from the economist and demographer Thomas Malthus, to the effect that humans—and, Darwin hypothesized, animals in general—tend to produce a surplus of offspring despite limited resources. Providing the missing piece of the puzzle, Darwin noted that this made possible an *evolution* of species, such that, given natural intra-species differences as a result of the offspring never being perfect copies of their ancestors, only those most fit for survival under the conditions provided will get the chance to reproduce their kind, which, over time, will give rise to adaptations as a result of a natural selection.

This is relevant to the matters at hand since such a remarkable synthesis of methods and data (the results of which were highly controversial) only are possible given an extraordinary attention to methodological considerations, requiring not only an ability to (*a*) identify methods correctly (and, hence, demarcate the relevant from the irrelevant factors) and (*b*) evaluate their robustness successfully—i.e., determine the extent to which they may be applied across experimental and theoretical domains—but also (*c*) apply or revise them within the new domain. And only given an exceptional sensitivity to the relevant aspects of methods can lessons from such diverse fields as geology, economy, and demography be successfully applied in biology. Darwin was, most likely, fully aware of this, and, upon returning from his journey on the Beagle, spent twenty three years deliberating over his empirical data, and the conclusions that could be inferred from them, before publishing his results in 1859, probably being highly sensitive not only to the epistemic qualities of the methods used, but also to the mental types pertaining to which he could trust his perceptual apparatus (e.g., in observing biological phenomena over extended periods of time under varying circumstances), or perform the required reasoning task (e.g., in inference).[25]

[25] Mason (1962, pp. 416-417).

By way of conclusion, successful science involves inquirers that are on-point token typers of the mental as well as of the methodological. In other words, within successful science, there are people who are good token typers with respect to the very domain of the generality problem—belief-forming processes and methods—and that, as such, solve generality problems on a daily basis. Furthermore, the fact that science constantly solves generality problems provides us with the material to solve the problem of vacuity. From the theoretical end, reliabilism suggests that justification is a matter of reliable belief-formation. From the practical end, science fills in the details as for exactly *what* types individual belief tokens fall under. However, this is not done by way of anything like a full-fledged theory of determination (again, scientists are notoriously bad at describing their own practice), but by virtue of demonstrably on-point capacities to distinguish relevant from irrelevant types on a case-by-case basis, as illustrated by the above examples from the history of science. Together, science and reliabilism, thereby, solve the problem of vacuity.

But what about the problem of testability? Remember, the problem of testability was the problem that philosophers might arbitrarily type tokens in a way that favors their particular theory. As I framed the problem above, it indicated a need for an adjudicator who can settle disputes regarding the typing of the mental and methodological, in cases where people's categorizations differ in ways that (*a*) makes a difference as for the evaluation of theories of justification, and (*b*) cannot simply be resolved through discussion or deliberation. In light of the recognitional capacities called attention to above, and the fact that science constitutes one of our most successful and fruitful intellectual endeavors, I suggest that science can be expected to constitute a very promising—if not the *most* promising—candidate for successfully adjudicating relevancy claims. That is, in cases where there is actually a genuine dispute as for what type a particular token belongs to (and I suspect that such disputes are somewhat less common than the proponents of the generality problem seem to think),

science may step in and use its recognitional capacities to either narrow down the live options or, if that does not solve the dispute, actually pick out the relevant type, understood either as a uniquely correct type, or a type that is part of a set of types that all serve the purpose of accurate prediction and experimental success to a sufficient degree. Clearly, this is *not* to say that scientists always coincide in their categorization claims. However, it *is* to say that the scientists behind successful science will be able to reach verdicts determinate enough to solve most relevancy disputes, so as to ensure that philosopher's relevancy claims are not just arbitrarily connected to the theory one happens to favor—and this is all the adjudication that we need. This solves the problem of testability, and concludes my solution to the generality problem.

7.7. Implications and Plan for Implementation

Now, let us turn to the epistemological implications of RC. To understand these implications, we first need to remember that unconscious heuristics tend to pertain to involuntary, often hard-wired processes, while reasoning strategies pertain to choices over which we may (and often do) deliberate. This has important implications for justificatory discourse. In one sense, whether or not I am justified in trusting, say, my perceptual beliefs might be something that I can not, in any helpful sense, do anything about, since it pertains to the workings of my perceptual apparatus, the make up of which I can not (on pain of invasive surgery) adjust. However, given information about how I am thus hard-wired may enable me to design my reasoning strategies so as to utilize or work around whatever sub-optimal aspects of my perceptual apparatus that I might come by (and, in the simple perceptual case, perhaps simply start wearing glasses or hearing aid).

This generalizes to all unconscious processes and provides an interesting point of departure for guidance directed epistemology. Against the background of the empirical work done on the actual track-records of lower order and largely automatic mental processes

and heuristics, such epistemology plays a role in the development of reasoning strategies that specify ways to reason—particularly in relation to the collection and processing of evidence—that tend to be effective, taking into account the imperfections as well as adaptive success of lower-order belief-forming processes and heuristics. This focus on reasoning strategies is also motivated by the fact that the type of question facing the typical epistemic inquirer, arguably, is not so much (the overly philosophical) "How do I act in accordance with my epistemic duties?" or "Do I have any good reasons to think that any of my beliefs about the world are true?" as the following more straightforward question, or rather question *schema*:

(Q3) Is it the case that *p*?

This brings us to my plan for implementing the reconstructed concept of justification: Taking (Q3)—as well as its tense modified versions "*Was* it the case that *p*?" and "*Will* it be the case that *p*?"—as one of the most central question schema of epistemic inquiry, the next chapter will discuss empirical research relevant to understanding the way we reason and what this research may teach us about how we could reason better.

As it happens, such empirical work currently makes up a flourishing discipline, perhaps most pertinently within cognitive psychology and, in particular, within the influential heuristics and biases program initiated by Daniel Kahneman and Amos Tversky[26] in the 1970's, as well as the more recent fast and frugal heuristics program of Gerd Gigerenzer and the ABC Research Group.[27] Taking this work into account turns our task of constructively analyzing justification into an explicitly empirical one, since it shifts the attention from pure concept-oriented philosophy towards the scientific studies that tell us

---

[26] See Kahneman, Slovic, and Tversky (1982) for a collection of some classical papers within the heuristics and biases tradition, and Gilovich, Griffin, and Kahneman (2002) for a more recent collection of essays.

[27] See, e.g., Gigerenzer and Selten (2001) and Gigerenzer *et al.* (1999).

what does, in fact, satisfy the above criterion for effective heuristics—something that can not be settled on a purely conceptual basis. Focusing on the development of sound reasoning strategies, thereby, also enables us to turn the analysis of justification away from detached armchair speculations to hands-on prescriptive epistemology, explicitly directed not only at studying but also at aiding epistemic inquiry in providing material for identifying the very strategies and processes that will serve us well in the pursuit of significant truths in naturalistic settings.

7.8. CONCLUSION

The present chapter started out with the quite pervasive idea that being justified pertains to having good reasons for taking a (set of) proposition(s) to be true—an idea that we then spelled out in three phases. First, we returned to points made in chapters 2 and 3 about the more specific goal of attaining and maintaining true belief in significant matters. Second, we ventured into the details of the specific *purpose* that JUSTIFICATION fulfills in relation to this goal, concluding that the purpose of JUSTIFICATION is to flag appropriate sources of information, as in sources that tend to produce true belief. Third, and in light of the failures of Deontologism and IBAI reviewed in chapters 5 and 6, it was argued that the most plausible way to understand this appropriateness was in terms of *effective heuristics*, designating justification conferring processes on a continuum stretching from unconscious belief-forming heuristics to conscious reasoning strategies, effective in so far as they strike a good balance between generating a lot of true belief (power) and generating a majority of true belief (reliability).

We also noted that one important—and in my opinion extremely welcome—consequence of such a concept of justification was that it classifies justification as a perfectly natural phenomenon, pertaining to the actual and possible track-records of belief-forming mechanisms. It was also noted that this, furthermore, has very intriguing methodological implications, in that it enables us to paint a more

positive picture than the one resulting from the largely negative task that has been preoccupying us in the course of rejecting candidate analyses of justification. As I will argue in the next chapter, the way to paint this picture is to focus on the pursuit of significant truths and, against the background of what cognitive science and psychology may teach us about the way we reason, to develop reasoning strategies that, enlightened by psychological research on our limits and potentials, may improve our cognitive outlooks and take us beyond the potentially bleak implications for human rationality and justification that some of the results may seem to motivate at first glance.[28]

[28] See, e.g., Nisbett and Borgida (1975).

Chapter 8.
On the Improvement of Reasoning Strategies

Having reconstructed JUSTIFICATION in terms of effective heuristics, it is the task of the present chapter to put this concept to work, in providing hands-on advice for epistemic inquiry. This is an ambitious task, to say the least, perhaps *too* ambitious, some would say. As pointed out by Goldman in *Epistemology and Cognition*—a valiant attempt to provide a related evaluation of our basic cognitive processes—cognitive science is still "groping its way toward the identification of basic processes."[1] This is, unfortunately, as true now as it was in 1986. And any truly naturalistic epistemology will necessarily be limited exactly to the extent that science is still to figure out the relevant details. Consequently, there is a respect in which naturalistic epistemology commits itself to contingency in result and conditionality in formulation.

Still, this is no excuse for the epistemologist to not get her hands dirty. Even though parts of her inquiry will have to remain tentative, it may still provide an idea of what a more satisfactory account would look like, if the relevant empirical details were to be filled

[1] Goldman (1986, p. 181).

in. The following is my attempt to outline such an account and I would like to preface it by pointing out some similarities as well as dissimilarities with two related attempts. First, Goldman's *Epistemology and Cognition*. Like Goldman, I am interested in the contribution that cognitive science can make to epistemology and, in particular, to the analysis of justification. However, Goldman is mainly concerned with a theoretical evaluation of the epistemic properties of basic psychological *processes*, or what I have referred to as unconscious heuristics.[2] My project is, in a sense, complimentary to his, in that I will mainly be concerned with the morals that can be pulled out of cognitive science for the evaluation and improvement of reasoning strategies. Furthermore, and unlike Goldman, I am explicitly interested not just in a theoretical evaluation but also in the construction of explicitly *regulative* strategies, designed to not only inform epistemology but also guide epistemic inquiry.

In this respect, my project is more in line with Bishop and Trout's, as laid out in their intriguing book *Epistemology and the Psychology of Human Judgment*.[3] In a critique of what they refer to as Standard Analytic Epistemology—defined as a methodology not too different from DCA (see §1.3 above)—they suggest that epistemology's mission should be to provide normative recommendations about how we ought to reason. The way to do this, they argue, is *not* by way of the analysis of epistemic concepts, but rather through what they call *ameliorative psychology*, which is a branch of psychology dealing with the construction of successful reasoning strategies on the basis of testable results. Bishop and Trout are particularly interested in the evaluation and construction of so-called *statistical predication rules* (SPRs), which are rules or algorithms yielding predictions on the basis of a specified set of cues and a (preferably simple) formula for combining these cues. We will have more to say about such rules in §8.2 below.

However, Bishop and Trout explicitly distance themselves from the idea that their normative framework presents anything like a

---

[2] See Goldman (1986, p. 184).

[3] See Bishop and Trout (2004).

theory of *justification*.[4] Justification, they claim, is a property of individual belief tokens and, hence, pertains to something which ameliorative psychology—occupied with the assessment and development of reasoning strategies—does not dwell on. As defined, or rather reconstructed, above, however, there is no such opposition on my account. Justification, as construed here, is primarily a property of epistemic agents believing propositions, and only derivatively of the belief tokens themselves. Furthermore, and as we saw in the previous chapter, what it is for an agent to be justified is intimately connected to—indeed, a function of—effective (i.e., reliable and powerful) heuristics, which, on the conscious side, corresponds to sound reasoning strategies. In this respect, I see a continuity between naturalistic and traditional epistemology that Bishop and Trout denies.

Again, the claim is not that this framework unveils our ordinary concept of justification. In that respect, I wholeheartedly agree with Bishop and Trout that theorizing about reasoning strategies is independent of questions of justification, as traditionally construed. Consequently, the claim that degree of justification is determined by the epistemic qualities of the heuristics underlying the beliefs held to be true by an agent, should not be expected to mesh with our categorization intuitions in all cases. To have our account of justification and our intuitions mesh thus has not been a desideratum of our analysis, for reasons that were discussed in the first part of the study. Granted, our intuitions reveal the outlines (be they fuzzy or not) of our concepts and an exhaustive analysis would, indeed, provide an exact and in all cases faithful story about these contours and, hence, not admit any conflicting intuitions. As argued in chapter 3, however, it is possible to question the very business of providing exhaustive analyses, since there is no guarantee that our current concepts—as unveiled

[4] See Bishop and Trout (2004, p. 54). Although, see Bishop (2000, pp. 203-205), where Michael Bishop seems to defend (or at least consider) a revision of what it is to be epistemically responsible—one component of which is justification—in terms of using reliable reasoning strategies.

through conceptual analysis—serve us well, i.e., provide us with good tools in the attainment of our epistemic goals.

For this reason, I argued that the operative question for epistemology should be whether the epistemic concepts and norms that we employ could be improved so as to enable us to reach our epistemic goals to a greater degree. Consequently, our current epistemic repertoire is primarily interesting to the extent that it informs us of a starting point of potential improvement. This, however, does not require an exhaustive analysis—it only requires what I, in chapter 4, introduced as Meaning Analysis. And as we saw in chapters 5 and 6, there is even evidence to the effect that, if "our current concepts" look anything like what epistemologists have supposed, they do, as a matter of fact, *not* serve us very well. This motivated reconstruction, in accordance with the recommendations of the ameliorative methodology defended in the first part of the study.

The reconstruction in question was expounded in the previous chapter. *Qua* reconstruction, the idea that justification consists in effective heuristics underlying the beliefs held true by an agent does not constitute an "analysis" of our concept of justification, in any traditional sense of the word, and is, hence, not susceptible to conflicting categorization intuitions. Our intuitions might be ever so internalist and maybe even deontological. Be that as it may—if the arguments put forward in chapters 5 and 6 are at all plausible, these intuitions, nevertheless, supply us with largely useless theories about justification, and should, as such, carry no more weight than folk intuitions about physical matter should to informed physicists. In this sense, the reconstructed concept of justification that I propose corresponds not to what we *do* mean but to what we *should* mean by justification. The "should" in question is an explicitly instrumentalist one. The idea is that, in so far as we subscribe to the truth-directed, epistemic goals that I have been claiming that we do, we are better served by using the reconstructed concept of justification than any of the concepts that tie justification to epistemic duties or introspective scrutiny, for the simple reason that the former will enable us to reach those goals to a greater degree than the latter two.

As we saw in chapter 2, this is also the proper locus of epistemic normativity in the case of justification—justification is valuable in so far as it is aligned with our epistemic goals, and our concept of justification is good in so far as it provides a useful tool in the attainment of these goals. In other words, the normativity of justification is inherently connected to our epistemic goals. If it is objected that this does not yield pure, *categorical* normativity, I would reply that it is unclear to me why epistemology—construed as a discipline thoroughly interested in not only describing epistemic inquiry, but also providing tools for epistemic inquirers that may aid her in attaining her goals—should be interested in anything but impure normativity, which, interestingly enough, seems to be exactly the kind of normativity that speaks to the needs, desires and typical situations of epistemic inquirers.

This brings us to the proper mission of this chapter, i.e., that of investigating the generally normative virtues of, in some cases, very basic heuristics in order to formulate action-guiding normative reasoning strategies, identifying ways in which the reconstructed concept of justification could be put to use in the creation of hands-on advice for epistemic inquiry. The plan is as follows: I will pull out three themes from cognitive psychological research on the way we tend to reason—including what can best be described as annoying cognitive tics, impressively adaptive heuristics, and several things in between—and investigate the extent to which it is possible to pull out concrete strategies for improving epistemic conduct in typical epistemic situations, using the reconstructed concept of justification as my normative framework. Focusing on three versions of question schema (Q3) above, i.e., "Is it the case that *p*?", the three themes that are to be discussed are *prediction*, i.e., reasoning about what *will* be the case, *diagnosis*, i.e., reasoning about what *is* the case, and *retention*, i.e., reasoning about what *was* the case.

Before doing this, however, I would like to spell out and, hopefully, appease some worries regarding the extent to which the project to be undertaken presupposes a far-reaching optimism as for our abilities to change the way we tend reason.

8.1. Reasons for Moderate Optimism

Any ameliorative approach to the ways in which we reason has to steer clear not only of the Scylla of assuming a too optimistic picture of our ability to change our epistemic ways, yielding a largely unrealistic and useless epistemology, but also the Charybdis of assuming a too pessimistic picture, rendering the resulting epistemology uninteresting and irrelevant. In this section, I will try to say something about how my own project steers clear of both.[5]

The most optimistic version of an ameliorative project would assume *perfect internalization*. By perfect internalization I mean the ability to take ameliorative suggestions to heart and let them permeate one's epistemic dispositions, as manifested in everything from everyday reasoning to sophisticated, scientific inquiry. However, cognitive psychology gives us little reason to be that optimistic. We have already seen how we tend to underestimate the extent to which we suffer from cognitive biases, despite the fact that every single one of us has substantial evidence to the effect that she is no different from anyone else on this score. And unfortunately, this tendency to over-rate our capabilities and ourselves is not restricted to the occasional anchoring or attribution error, but pervades our overall sense of who we are and what we are capable of doing.[6] Adam Elga sums up the psychological research as follows:

> It turns out that people have inflated views of their own abilities and prospects. People (nondepressed people, at least) rate themselves better—friendlier, more likely to have gifted children, more in control of their lives, more likely to quickly recover form illness, less likely to get ill in the first place, better leaders, and better drivers—than they really are. And that's just

---

[5] Thanks to Hilary Kornblith, Kirk Michaelian, and Jeremy Cushing for pressing me on this point.

[6] See, e.g., Taylor and Brown (1988, 1994).

> the beginning. There is a great deal of work documenting the persistent and widespread positive illusions (about themselves) to which people are subject.[7]

As Elga also points out, depressed people have been found to have more accurate self-evaluations than non-depressed people. In fact, there is even reason to believe that there is a strong connection between the presence of positive illusions and increased happiness, motivation, persistence, and the ability for productive work.[8]

In light of this, perfect internalization might not only be *unfounded*—the mere deliverance that certain ameliorative measures can be taken simply does not seem to have that great of an impact on us—but also *undesirable*, given the connection between positive illusion and happiness. So, let us turn to the second extreme, which I would like to refer to as *dismal incorrigibility*. If we assume dismal incorrigibility, we are assuming that our epistemic ways are so set in stone that no attempt to improve them has any likelihood whatsoever of succeeding. Clearly, given such an assumption, there would be no reason to engage in any form of amelioration; it would be doomed to fail already from the start. Luckily, however, one does not need to look very far to find evidence refuting dismal incorrigibility. For one thing, methodological considerations within the sciences has, clearly, enabled us to not only identify but also implement improved ways of reasoning, for example through formal logics as well as through applied mathematics, especially in the form of probability theory and statistics. So, by *modus tollens*, dismal incorrigibility cannot be true.

Having thus steered clear of Scylla and Charybdis, it remains to say something more substantive about the exact route to be taken in between the two—a route I will refer to as *moderate optimism*. First, moderate optimism acknowledges the empirical data ignored by perfect internalization. As we have already seen, we suffer from a great variety of cognitive biases, even some of which involve the common

---

[7] Elga (2005, p. 117).

[8] See Taylor and Brown (1988).

conviction that we do not. However, given extensive research on these biases, and the conditions under which they may be more or less severe, we may alleviate them in situations where it is important that our epistemic goals be satisfied. The qualification in terms of such situations is important since one crucial difference between moderate optimism and perfect internalization is that the former does *not* assume that amelioration can take place across the board. In fact, as already noted, the aforementioned research on the connection between positive illusions and happiness might even suggest that across-the-board amelioration would be directly undesirable, even if attainable.

However, this is not to deny that there are situations in which positive illusions are harmful and accurate reasoning is important. Medical diagnostics is a good example, and also one that we will discuss at length below. Nor is it to deny that nothing can be done to make accurate reasoning a more prevalent feature in such situations. For example, it will be suggested in the following that even a brief training in statistics may lead to significant improvement in predictive reasoning. Furthermore, it will be suggested that framing such statistical data in terms of frequencies rather than in terms of probabilities may decrease the chances of errors to an even greater extent. Finally, I will look into some results indicating that taking care to frame one's data in terms that enable one to elaborate on them in depth, and then maximize the lag between the periods of elaboration, will significantly increase memory power, which, given accurate input, may increase the accuracy of our beliefs.

The empirical assumptions underlying my interest in these results are that (*a*) we often fail to live up to what has traditionally been considered ideal rationality (roughly defined in terms of conduct in complete accordance with probability theory, principles of maximizing expected utility, etc.), (*b*) it is possible to pin-point quite well in what ways we fail to do so since it is, after all, the case that (*c*) we do sometimes engage in good (i.e., roughly, truth-conducive) reasoning and, finally, that (*d*) pin-pointing the ways in which we fail to live up to ideal rationality may enable us to device reasoning strategies that

bring out our good rather than bad cognitive tendencies. I take it that (*a*) is fairly uncontroversial. So is (*b*), at least among people that have any interest in an empirical study of human reasoning. And to the extent that one accepts (*b*), one seems committed to (*c*)—after all, if one was convinced that we suffer from several cognitive biases but did not think that we, at least occasionally, engaged in sound reasoning, the only reasonable strategy would be to shun honest inquiry altogether (at least if honest inquiry implies an honest inquiry into the truth).

However, among the four, it is (*d*) that would motivate the particularly *ameliorative* mission that I will undertake in this chapter, and it is also that very assumption that, ultimately, motivates moderate optimism. This, however, is not to say that moderate optimism rests on an article of faith. Rather, I take it that the results to be considered, and the fact that they indicate ways in which human reasoning not only *can*, but, in fact, *has* been significantly improved, within such diverse areas as prediction, diagnosis, and retention, all provide reasons—indeed, straightforwardly *empirical* reasons—for embracing a moderate optimism. So, without further ado, let us turn to our first area of investigation: prediction.

8.2. Prediction and Statistical Reasoning

Over the last fifty years, psychologists have amassed a substantial body of data indicating that, in many clinical situations, the most accurate predictions are made not by relying on clinical judgments of experts, but by using so-called linear prediction models. A linear prediction model (I will henceforth leave out the qualifier "prediction") is, in essence, a simple weigh-and-add equation of the following form:

$$V = w_1c_1 + w_2c_2 + \ldots w_nc_n.$$

V here represents the predicted value of a target property (i.e., the property to be predicted), c_1 through c_n a set of cues, and w_1 through w_n the weights assigned to those cues. The first evidence to the effect

of the superiority of such surprisingly simple models came when Paul Meehl famously reviewed 22 studies comparing the clinical judgments of expert psychologists and psychiatrists with linear models based on nothing but the empirical data on the relevant events.[9] The results were stunning; in all studies, the linear models either performed equally well or outperformed the expert clinicians.

Following up on Meehl's study twelve years later, Jack Sawyer reviewed 45 studies comparing clinical and statistical predictions via linear models.[10] Again, not in a single study were the former superior to the latter. As if this was not enough, Sawyer even included two studies in which the clinicians had access to *more* information than was fed into the linear models, only to find that the clinicians under those conditions, in fact, performed even *worse*. As noted more recently by Robyn Dawes, David Faust, and Meehl, these and similar research outcomes have since been widely replicated, to the extent that there are now "nearly 100 comparative studies in the social sciences" such that, "[i]n virtually every one of these studies, the actuarial [i.e., statistical] method has equaled or surpassed the clinical method, sometimes slightly and sometimes substantially."[11]

Why are linear models so successful? To answer this question we need to make a distinction between two types of linear models that have figured in the literature: proper and improper. Proper linear models are linear models with weights devised to maximize the fit between the component cues and the target property. This is not the case for improper linear models, the most prominent forms of which are random linear models, unit weight models, and bootstrapping models. In the case of random linear models the weights are assigned random values, while unit weight models are such that every cues is assigned equal weight. Bootstrapping models are devised so as to not directly mirror the target property but rather mimic the prediction of a subject (e.g., a clinician) predicting that property.

---

[9] See Meehl (1954).

[10] See Sawyer (1966).

[11] Dawes, Faust, and Meehl (2002, p. 719).

Now, if the aforementioned superiority of linear models over clinical judgments was restricted to proper models, explaining their success might not have been that big of a challenge. It could have been explained, at least in part, with recourse to how the employment of such models minimizes the influence of performance errors pertaining to limitations in memory, attention and computational capabilities on part of the person doing the prediction.[12] As noted by Dawes, Faust and Meehl, such factors as "fatigue, recent experience, or seemingly minor changes in the ordering of information or in the conceptualization of the case or task can produce random fluctuations in judgment [which] decreases judgmental reliability and hence accuracy."[13] In contrast, Hastie and Dawes points out, a proper linear model "uses valid, independent information from as many cues as convey such information, is 'calibrated' to the ranges of values on all the variables available in the situation, and operates relentlessly and consistently."[14] However, one of the most astounding results to come out of the research on linear models is that clinicians generally are outperformed not only by proper but also by improper models.[15] How do we explain *this*?

Granted, the above explanation in terms of performance errors may provide part of the explanation of the relative success of bootstrapping models over the clinicians they are devised to mimic. If it can be assumed that the bootstrapping model latches on to the heuristic that the clinician is using, but fails to use *perfectly* due to limi-

---

[12] For a more complex explanation of the success of proper linear models, invoking assumptions about the prevalence of ordinal or monotone (rather than crossed) causal interactions in nature and the ease with which such causal relationships are approximated by linear models, see Hastie and Dawes (2001, pp. 58-62).

[13] Dawes, Faust, and Meehl (2002, p. 724).

[14] Hastie and Dawes (2001, p. 62).

[15] See Goldberg (1970) for a classical study on bootstrapping models, Dawes and Corrigan (1974) on random linear models, and Howard and Dawes (1976) on unit weight models.

tations in memory, attention, and computational capabilities, it should come as no surprise that the model often outperforms the clinician. Still, this explanation, clearly, does not generalize to random linear and unit weight models. Instead, one of the most popular explanations for the success of improper models rather refers to the so-called *flat maximum principle*.[16] What this principle states is that, under certain conditions, any linear model (be it a proper or improper) will perform roughly as well as any other. These conditions are as follows:

(*a*) The signs of the coefficients are right.
(*b*) The cues used in the model are fairly predictive yet somewhat redundant.
(*c*) The prediction pertains to a difficult problem in the specific sense that no proper model will be especially reliable.

As one might reasonably suspect, many social prediction problems—Which business venture will maximize share-holder value? Which prospective graduate student will perform best if admitted? Which convict will fall back into crime if released from prison?—pertain to fairly difficult problems, in the sense that no predictive model will be that much more reliable than any other, and most predictive cues that we come by in these contexts will be fairly predictive yet somewhat redundant. And if so, the only thing we have to do is get the signs of the coefficients right. This explains why improper linear models often tend to be as reliable as proper linear models. Furthermore, since proper linear models tend to be just as or more reliable than clinicians, we thus also have an explanation of why improper linear models often outperform clinicians.

Even explanation aside, however, linear models demand our attention as manifestly very powerful predictive tools. As Meehl stated already in 1986:

[16] See, e.g., Einhorn and Hogarth (1975) and Lovie and Lovie (1986).

> There is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing 90 investigations, predicting everything from the outcome of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion, whatever theoretical differences may still be disputed.[17]

The conclusion, it seems, is fairly straightforward:

> *Clinical Prediction via Linear Models (CPLM)*
> When faced with a clinical prediction problem, ask the experts what cues and weights to use (unless you plan to employ a random or unit weight model) and then let a linear model combine the information from those cues to make the judgment.[18]

The rationale for such a strategy is, of course, that, if there is any truth to the research just discussed, other things being equal, clinical predictions made on the basis of linear models tend to be just as accurate as, and often more accurate than, those based on the actuarial judgments made by clinicians. This is a somewhat more modest version of Bishop and Trout's *Golden Rule of Predictive Modeling*, stating: "When based on the same evidence, the predictions of SPRs [i.e., statistical prediction rules] are at least as reliable, and are typically more reliable, than the predictions of human experts."[19]

Let us consider an example. Take a medical doctor who receives a patient complaining over pains that may indicate an ensuing heart attack. Being pressed for time and resources, the doctor needs

---

[17] Meehl (1986, pp. 373-374).

[18] Cf. Hastie and Dawes (2001, p. 58).

[19] Bishop and Trout (2004, p. 12).

to act fast and asks herself "Is it likely that this patient will suffer a heart attack anytime soon?" When answering this question, she should start by considering whether there are any applicable prediction models relevant to, in this case, classifying potential heart attack victims. As a matter of fact there is. Take, for example, the following very simple model, adapted from Breiman *et al.* working with binary yes/no values rather than real numbers, and where a high-risk patient is supposed to designate someone likely to suffer from a heart attack soon, and a low-risk patient someone who is not:

> Is the minimum systolic blood pressure over the initial 24 hour period > 91?
>
> If no, the patient is high risk.
>
> If yes, is the patient's age > 62.5?
>
> If no, the patient is high risk.
>
> If yes, is sinus tachycardia present?
>
> If no, the patient is low risk.
>
> If yes, the patient is high risk.[20]

It should be noted that, when applying this method, it is important the doctor assigns a greater weight to the recommendations derived from the method than to her own clinical expertise. The reason is, of course, that, if there is any truth to the research surveyed above, prediction models are more likely to produce accurate predictions when not taken in conjunction with the supposedly less reliable judgment (relatively speaking) of the doctor going on her own clinical expertise.

Clearly, there is a need to further specify exactly what models to use in what situations. Although the task of providing such a specification goes far beyond the present study, I encourage the reader to look at the vast body of literature on the subject since Meehl's study, ranging from predicting loan and credit risks,[21] the chance of criminal recidivism,[22] violent behavior,[23] and the risk of

---

[20] See Breiman *et al.* (1993).

[21] See Stillwell *et al.* (1983).

recidivism,[22] violent behavior,[23] and the risk of SIDS[24] to the quality of red Bordeaux wine.[25]

Here, however, I will instead turn to another question: What about the cases in which there are no available models? To answer this question, we need to make a distinction between two aspects or stages of prediction by way of linear models. More specifically, I suggest that we understand prediction by way of linear models in terms of (*a*) the *selection* of a set of weighted cues and (*b*) the production of a prediction through the *combination* of these weighted cues. In the cases where there are models in place, we do no need to worry about the selection phase—the selection of cues is already done. So, in those cases, we can simply heed to CPLM and apply the model in question.

However, in the cases where there are no applicable models, we need to come up with new ones. In the following, I will focus on the crucial yet less discussed selection phase and demarcate conditions under which people select better rather than worse cues. After all, in so far as we want to engage in prediction through statistical reasoning, there is no getting around the fact that someone has to pick out the weighted cues and that there, in many high-risk situations, are obvious reasons for why we would want that selection to be made as sensibly as possible; sometimes, the cost of an inaccurate prediction is simply too high. I will assume that one important factor in making such selections is a sensitivity to statistical information. By way of illustration, let us start out by considering a scenario, described by Nisbett, Krantz, Jepson, and Kunda:

> Harold is the coach for a high school football team. One of his jobs is selecting new members of the varsity team. He says the following of his experience: "Every year we add 10 to 20 younger boys to the team on the basis of their performance at

[22] See Carroll *et al.* (1982).

[23] See Faust and Ziskin (1988).

[24] See Carpenter and Emory (1977).

[25] See Ashenfelter, Ashmore, and Lalonde (1995).

> the try-out practice. Usually, the staff and I are extremely excited about the potential of two or three kids—one who throws several brilliant passes or another who kicks several field goals from a remarkable distance. Unfortunately, most of these kids turn out to be only somewhat better than the rest."[26]

What psychologists have found is that, when faced with the question why Harold usually has to revise downward his opinion of players that he originally thought were brilliant, people tend to invoke a series of causal explanations—Did Harold's eagerness lead him to overestimate the brilliance of some players during the try-out? Did the selected players slack off in face of all the encouragement they may have gotten from Harold after the try-out?, etc.—while ignoring a very basic statistical phenomenon that is likely to occur in situations like these, namely *regression toward the mean*.

Consider the first sample at the try-out: a fairly large selection of high school students. Within this selection, some will be better football players than others. If being a good football player was a completely transparent property, Harold's job would be easy. One factor that makes it harder, however, is luck—good as well as bad. Sometimes good football players perform badly for no other reason than that they happened to be unlucky and sometimes bad football players perform well for no other reason than that they happened to be lucky. Maybe they just happened to stand at the right (or wrong) place at the right (or wrong) time; maybe a lucky (or unlucky) gust of wind happened to assist them in throwing that magnificent (or disastrous) pass. In other words, luck works both ways. However, given that there are more un-exceptional than exceptional players (which seems to follow from the meaning of the terms), it is more likely that an exceptional performance was the result of an un-exceptional player who got lucky than that a poor performance was the result of an exceptional player who got unlucky.

---

[26] Nisbett *et al.* (2002, p. 523).

This is not to deny, of course, that any given player is equally likely to have good or bad luck. But now consider the second sample, i.e., the sample we have got after the try-out: a significantly smaller selection based on the performance at the try-out. If the above is true, picking out this sample on the basis of performance on the try-out makes good luck matter more than bad luck. More specifically, given the preponderance of un-exceptional players in the first sample and the equiprobability of good and bad luck for any given player, there are more opportunities for good luck to shift someone to the category of exceptional performance than for bad luck to shift someone to the category of unexceptional performance. Consequently, picking out a sample on the basis of exceptional performance at the try-out is more likely to reveal good luck than bad luck. However, given that good luck is a significantly less stable property than actual competence, this good luck cannot necessarily be counted on in the varsity team. Hence, it is to be expected that the performance of the new varsity team players will regress toward the mean, which, given the preponderance of un-exceptional players, unfortunately, means that they will perform worse than they did on the try-out.

In fact, regression effects occur in any scenario in which variables (here, performance at try-out, performance on the varsity team player) are not perfectly correlated. Furthermore, the above example may easily downplay the potential problem with neglecting regression toward the mean. Consider the following scenario from Kahneman, Slovic, and Tversky:

> In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improved on the next try. The instructors concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial, contrary to accepted psychological doctrine. This conclusion is unwarranted because of presence of regression toward the mean. As in other cases of repeated examination, an

> improvement will usually follow a poor performance and a deterioration will usually follow an outstanding performance, even if the instructor does not respond to the trainee's achievement on the first attempt. Because the instructors had praised their trainees after good landings and admonished them after poor ones, they reached the erroneous and potentially harmful conclusion that punishment is more effective than reward.[27]

By the same token: Within public policy, any intervention aimed at an unusual characteristic or a group that is very different from the average is likely to appear successful, while success often is nothing but an instance of regression to the mean. This is perhaps especially pertinent in public health interventions, which are often aimed at sudden increases in disease. Analogously, the phenomenon can lead to misinterpretation of results of tests, new treatments, as well as to a placebo effect in clinical practice, especially if participants in the studies are recruited on the basis of scoring highly on a symptom index. In other words, it should be beyond doubt that regression toward the mean is a phenomenon the ignorance of which affects issues of high significance.

How does this relate to the selection phase? Most likely, if Harry was more versed in statistics, he would not take the fact that certain boys performed extraordinarily well at the try-out to be such a powerful predictor of extraordinary performance at the varsity team, and, hence, would lover the weight assigned to that cue, while assigning a greater weight to long-term performance—a cue more sensitive to regression toward the mean. Similarly, if the flight instructors in the above example were more familiar with statistics, they might have been able to see through the tempting but most likely false explanation of improved performance in terms of admonishment and instead have predicted future performance on the basis of some more valid

---

[27] Kahneman, Slovic, and Tversky (1982, p. 10). This particular example is discussed in more detail in Kahneman and Tversky (1973).

cue, such as whether or not the pilots in question had undergone extensive and repeated training. And if public health interventions were conducted against the background of a better understanding of common, statistical phenomena, they would be less likely to invite the misinterpretation of research results, and, thereby, promote public health to a greater extent.

In other words, in order to pick out good cues, people need to be sensitive to statistically significant properties and phenomena. Here, however, we run up against a long and largely negative psychological tradition according to which people tend to be quite bad at reasoning in accordance with statistics and probability theory.[28] Rather than delving into this debate, however, I will (*a*) focus on what most researchers agree on, namely that there are conditions under which people do *not* tend to commit statistical fallacies, and then (*b*) say something about how these conditions may be generalized into a sensible reasoning strategy for cue selection.

Before looking into the specific conditions under which we are likely to engage in *correct* statistical reasoning, let us try to determine the general conditions under which we tend to reason statistically at all. "Statistical reasoning" is here to be understood as any reasoning about populations by way of samples and relations of frequency or probability (not necessarily involving any kind of numerical or formal representation), e.g., "You always forget to lock the door, so I'm sure you forgot to do so this morning too." Such reasoning should be contrasted with *causal* reasoning, which might very well invoke probabilities, but that typically only involves considerations about the causal relationships and interactions between individual entities or phenomena—e.g., "I overslept this morning and had to run to the bus. That's why I forgot to lock the door"—rather than populations at large, samples therein, and any statistical regularities that may hold between the two.

Now, according to Nisbett *et al*, there are three main factors conducive to statistical reasoning, the first one being *clarity of sample*

[28] See Kahneman, Slovic, and Tversky (1982).

space and sampling process.[29] In other words, people are more prone to use statistical reasoning (other things being equal) when pondering the chance of a coin falling heads or tails than whether to join this or that school, since the former case involves a clearer sample space (heads, tails) than the latter (good or bad relationship with the faculty and fellow students, good or bad campus, good or bad prospects for a career, etc.) and, hence, a more straightforward repeatability, which makes it easier to conceptualize one's observations as samples.

The second factor conducive to statistical reasoning is *recognition of the operation of chance factors*. For example, due to the salience of chance and, as a consequence, unpredictability in the production of an event such as coin tossing, and the much more subtle aspect of chance in the events following the choosing of one school over another, we are more prone to explain the outcomes in statistical terms in the former than the latter kind of case, rather than coming up with a causal explanation.

The third and, perhaps, most straightforward factor is *cultural prescriptions of statistical reasoning*. That is, people are more prone to use statistical reasoning within a given domain when statistical reasoning is the culturally prescribed way to think about events within that domain. Furthermore, Nisbett *et al.* found a connection between *experience* within an area and a tendency to use statistical reasoning in explaining events within that area. For example, when presenting the example of Harold the football coach to 157 University of Michigan students, subjects with an athletic team experience preferred a statistical explanation while a majority of those without any such experience preferred a non-statistical explanation. The same phenomenon was found in an analogous formulation of the scenario in terms of auditions, where most of the subject with acting experience preferred a statistical explanation while those without acting experience preferred a non-statistical explanation.[30]

---

[29] See Nisbett, Krantz, Jepson, and Kunda (2002).

[30] See Nisbett *et al.* (2002, p. 524).

Now, of course, a *proneness* to reason statistically is not, in itself, valuable unless one also tends to reason *correctly*. However, it is highly plausible to assume that training in statistics improves people's abilities to apply those principles correctly, since such training provides general skills that not only facilities identification of sample spaces and processes as well as the potential operation of chance, but also makes accessible and expands the repertoire of statistical rules. Hence, in a study on college students without training in statistics, graduate students with a fair amount of training, and PhD level scientists with several years of training, Krantz, Fong, and Nisbett not only found a correlation between training in statistics and a tendency to give statistical explanations (whether or not there was a statistical cue in the examples provided to the subjects) but also that only 10% of the statistical answers given by the first category were rated as good, while almost 80% of the answers from the third category were rated as good.[31]

So, consider the following reasoning strategy:

> *Selection via Statistical Training (SST)*
> Learn (more) statistics so that, when you face a prediction problem for which there is no established prediction model, you can not only identify as clearly as possible the relevant sample spaces and processes as well as the potential operation of chance, but also utilize this information in picking out statistically relevant cues.

In accordance with the research just surveyed, the rationale for this strategy is that learning statistics will serve the dual function of both (*a*) increasing the likelihood of you framing events in statistical terms and, hence, not neglecting relevant, statistical phenomena, and (*b*) providing you with the statistical tools necessary for picking out statistically relevant cues.

---

[31] Manuscript discussed in Nisbett *et al.* (2002).

So, what does it meant to "learn (more) statistics"? As noted by Bishop and Trout, when considering new reasoning strategies, we have to take into account *the start-up cost*—i.e., "the cost associated with adopting new reasoning strategies,"[32] pertaining to search, implementation, etc. And interestingly enough, another study by Fong, Krantz, and Nisbett suggests that even a very brief training in statistics can make a big improvement in people's statistical reasoning about everyday events.[33] In the study, subjects were given a training package, covering formal aspects of sampling and the law of large numbers, as well as demonstrating how sampling notions may be used as heuristic devises in modeling statistical problems. When weighed against a control group given no instructions, three experimental groups consisting of adults and high school students showed significant training effects, especially for those who had received a combination of sampling and modeling training. What is more, Fong, Krantz, and Nisbett showed that it made no difference whether the training utilized probabilistic cues, objective attribute problems, or subjective judgment problems, which, according to Nisbett *et al.*, suggests that "training on specific problems types can be readily abstracted to a degree sufficient for use on widely different problem types."[34]

In other words, the start-up cost of SST might not be as substantial as one might initially suspect, and even brief statistical training does, in fact, render people (*a*) more prone to reason in statistical terms and (*b*) less prone to commit various statistical fallacies than people who do *not* heed to SST. Consequently, heeding to SST in the selection of cues for linear models will, most likely, promote the production of true belief at a reasonable cost, and enable epistemic inquirers to construct efficient linear models, that then may be employed in accordance with CPLM. It should be noted, however, that this certainly is not to say that everyone should learn statistics (in accordance with SST) and always employ linear models in prediction

---

[32] Bishop and Trout (2005, p. 62).

[33] Manuscript discussed in Nisbett *et al.* (2002).

[34] Nisbett *et al.* (2002, p. 530).

(in accordance with CPLM). Such specific recommendations would have to be preceded by an evaluation of the situational costs of taking such measures, and how these costs relate to the possible detriments involved in making false predictions.

What is more interesting for our purposes, however, is the result we get when combining the above results with the reconstructed justificatory framework developed in the previous chapter:

> *S* is justified in believing that *p* to the extent that *p* is produced by effective heuristics, i.e., heuristics that, in all or a very wide sub-set of the epistemically relevant worlds, (*i*) generate a lot of true beliefs, and (*ii*) generate a set of belief that will contain a majority of true beliefs.

First, consider contexts with available prediction models. If there is any truth to the above research, we may infer the following:

> (8.2.1) Other things being equal, people basing their clinical predictions on linear models are just as justified, and often more justified, in the resulting beliefs than people basing their clinical predictions on the inferences of expert clinicians.

A couple of comments are at place. First, for reasons discussed above, justification requires not only effective reasoning strategies but also effective unconscious heuristics, i.e., underlying belief-forming processes. Naturally, employing linear prediction models rather than consulting clinicians directly does not, in itself, imply that the models are being employed for the right reasons, e.g., as a result of sound deductive or inductive reasoning, as opposed to guesswork, or wishful thinking. That is, it might very well be possible that someone chooses their strategies by way of frail strategy determining heuristics.

Consequently, as stated, (8.2.1) does not imply that you are always more justified in working with linear prediction models rather than consulting clinicians in predictive matters. For one thing, such a

claim would have to be made under the assumption that the strategy determining heuristics involved in choosing one reasoning strategy over another are, as a matter of fact, at least as effective as those underlying the choice of consulting a clinician. Consequently, a person employing a highly effective linear prediction model may still be less justified than a person consulting a clinician, if the strategy choice of the former was a result of, say, guesswork and the choice of the latter was a result of sound inductive reasoning.

For another—and this brings us to the second point—parity of strategy determining heuristics is not the only factor embedded in the "other things being equal" clause. Another such factor is the *power* of the reasoning strategy used, be it employing a linear model or whatever heuristic the clinician's inferences pertain to. Consequently, (8.2.1) is perfectly compatible with claiming that a person basing her predictions on the inferences of clinicians may be more justified than someone basing her predictions on inferences reached by way of a linear model, due to the former being the result of a more powerful reasoning strategy.

Next, let us turn to the scenarios in which there are no applicable prediction models. Combining the above results in relation to SST with our justificatory framework, we get the following:

> (8.2.2) Other things being equal, people who heed to SST (or any stronger variant thereof) in the selection of cues, tend to be just as justified, and often more justified, in the beliefs resulting from the employment of the forthcoming predictive model than people that do not do so.

As above, this is certainly not to say that everyone should learn statistics (in accordance with SST). Such specific recommendations would have to be preceded by an evaluation of the situational costs of taking such measures, and how these costs relate to the possible detriments involved in constructing unhelpful models. However, all that is required for the strategies suggested here to be at all relevant is that

there are numerous situations in which implementing demonstrably effective strategies is worth the cost. Needless to say, such situations exist. Hence, the centrality of teaching statistics to people of high-risk judgment professions, the prevalence of statistical reasoning within many scientific areas, and the significance that we tend to bestow research on the extent to which we do or do not fall prey to various statistical fallacies. That being said, we will, in the following section, look into yet another aspect of statistical reasoning as a means to attaining our epistemic goals and, in particular, how certain ways of framing statistical data may result in substantially more tractable reasoning strategies than others.

8.3. Base Rate Neglect in Diagnosis

I will now turn to so-called base rate neglect in diagnosis. Base rates are prior probabilities, frequencies or proportions, for example regarding the prevalence or exceptionality of a particular phenomenon in a given sample. Base rate neglect typically occurs when subjects are reasoning from symptoms (i.e., effects) to causes in order to reach a conditional probability (the probability of the cause given the symptoms) and, in doing so, simply equates this probability with its inverse (the probability of the symptoms given the cause). For example, consider a hypothetical case where a test *T* tests for condition *C* with an accuracy of 80%. Now, say that a subject *S* has been tested for *C* via test *T* and the test comes out positive. What is the probability that *S* has *C*? Faced with this kind of question, people tend to judge that the probability is 80%, thereby equating the conditional probability

$$P((S \text{ has } C) \mid (S \text{ tests positive for } C \text{ on } T)),$$

with its inverse

$$P((S \text{ tests positive for } C \text{ on } T) \mid (S \text{ has } C)).$$

The mistake in doing so may be brought out by a simple analogy: The probability that Amy is pregnant given that she has had sex is, clearly, not equal to the probability that she has had sex given that she is pregnant.[35]

So, what is the correct way to approach these kinds of questions? If we turn to probability theory, the standard approach invokes Bayes' Theorem, stating that

$$(8.3.1) \quad P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(A) \times P(B \mid A) + P(\neg A) \times (P(B) \mid \neg A))},$$

where A represents the (alleged) cause and B the symptom or effect. As this theorem makes clear, the conditional probability of A given B is *not* equal to the probability of B given A, but instead intimately tied to the base rate, i.e., to the prior probability of A and the prior probability of *not* A.

Now, given that Bayes' Theorem is something you typically do not encounter (let alone utilize) until you have taken one or two classes in statistics or probability theory, it should come as no surprise that people in general do not automatically reason in accordance with Bayes' Theorem and, hence, tend to ignore base rates. However, further studies indicate even people that can be expected to be fairly well versed in statistical reasoning in general and Bayes' Theorem in particular fail to use this theorem in diagnostics. In a study on the faculty and staff at Harvard Medical School on a diagnosis problem similar to the one above, only 18% provided the answer given by Bayes' Theorem.[36] Similarly, David Eddy asked 100 practicing physicians to estimate the (posterior) probability of cancer given a positive result, only to find that 95 of them estimated it to be between 70% and 80%, rather than the 7.8% yielded by Bayes' Theorem.[37]

What are the normative implications of this? At the face of it, this might seem to lend some credibility to Tversky and Kahneman's

---

[35] This analogy is borrowed from Bishop and Trout (2005, p. 139).

[36] See Casscells, Schoenberger, and Grayboys (1978).

[37] See Eddy (1982).

claim that "In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not a Bayesian at all."[38] And given that Bayesian reasoning provides a powerful tool in the attainment of truth—indeed, in such highly significant matters as medical diagnostics—that would clearly be bad news for the epistemic inquirer (not to mention the patient). However, there is reason to believe that the aforementioned failures reviewed by psychologists are artifacts of a particular way of presenting probabilistic information rather than evidence that we are inherently bad at reasoning in statistical terms.

Gerd Gigerenzer and Ulrich Hoffrage has argued that any claim against the human instantiation of a cognitive *algorithm*—be it a Bayesian one or not—is impossible to evaluate unless the information *format* on which it is designed to operate is also specified.[39] More specifically, they hypothesize that the mind has not evolved to handle statistical inferences in just any format and, in particular, that the mind is more prone to reason in terms of frequencies than probabilities. This, however, does not imply that the human mind is incapable of reasoning in Bayesian terms, but rather that whether or not they are able to do so will be strongly dependent on the particular format in which the problem she is confronted with is framed.

So, consider the following hypothetical scenario:

> Imagine an old, experienced physician in an illiterate society. She has no books or statistical surveys and therefore must rely solely on her experience. Her people have been afflicted by a previously unknown and severe disease. Fortunately, the physician has discovered a symptom that signals the disease, although not with certainty. In her lifetime, she has seen 1,000 people, 10 of whom had the disease. Of those 10, 8 showed the symptom; of the 990 not afflicted, 95 did. Now a new pa-

[38] Kahneman and Tversky (1972, p. 450).

[39] See Gigerenzer and Hoffrage (1995).

> tient appears. He has the symptom. What is the probability that he actually has the disease?[40]

Now, to calculate the probability, the physician may utilize either of two mathematically equivalent formats. Either, she may use Bayes' Theorem, in accordance with (8.3.1) above. Or she may go with what Gigerenzer and Hoffrage calls "natural sampling," in which case all the physician needs is the number of cases that had both the symptom and the disease (i.e., 8) and the total number of symptom cases (8+95) and then solve the following equation:

$$(8.3.2) \quad P(A \mid B) = \frac{B \,\&\, A}{B \,\&\, A + B \,\&\, \neg A} = \frac{8}{8+95},$$

where $B \,\&\, A$ is the number of cases with symptom and disease and $B \,\&\, \neg A$ the number of cases with the symptom *lacking* the disease (i.e., false positives).

Clearly, (8.3.2) is more transparent and tractable than (8.3.1), but why? For one thing, it involves fewer and seemingly less cognitively demanding operations. But remember that Gigerenzer and Hoffrage's hypothesis is stronger than this; they want to claim that the tractability of equations like (8.3.2) pertained not primarily to them being less computationally demanding but to their frequentist format, and that this format is more suited than its probabilistic equivalent to fit our cognitive architecture. If that is so, however, it is not sufficient to contrast a complicated Bayesian format with a simple instance of a frequentist equivalent. More specifically, if Gigerenzer and Hoffrage's hypothesis is true, we should expect to get the following experimental result: If subjects are presented with a diagnosis problem, they will have a significantly greater success in approximating the Bayesian answer if the problem is framed in terms of frequencies rather than probabilities, even in cases where the computations involved can be expected to be equally demanding.

---

[40] Gigerenzer and Hoffrage (1995, pp. 686-687).

In fact, this is exactly what Gigerenzer and Hoffrage found. When presented with a problem framed in probability terms, only 16% got the Bayesian answer. However, when presented with the same problem framed in frequentist terms and involving an equally complex or parsimonious algorithm, the percentage rose to 46%.[41] Similar results have been found by Leda Cosmides and John Tooby.[42] Presenting 50 Stanford students with questions similar to the ones presented by Casscells *et al.* to a group of faculty, staff, and fourth-year students at Harvard Medical School, 76% of the participants in Cosmides and Tooby's study gave the correct Bayesian answer—as compared to 18% in Casscells *et al.*'s original study.[43] And what was the difference? Cosmides and Tooby framed their questions in terms of frequencies, not probabilities.

Interestingly enough, the results seem to generalize to other "fallacies" as well. Take the following story, for example:

> Ben is an alcoholic tennis star who starts drinking a fifth a day on his 25th birthday. Which of the following two scenarios is more likely?
>
> (1) Ben wins a major tournament shortly before his 26th birthday.
> (2) Ben joins Alcoholic Anonymous, quits drinking, and wins a major tournament shortly before his 26th birthday.[44]

Despite the fact that scenario (2) *cannot* be more likely than scenario (1), for the simple reason that the probability of a conjunction of events cannot be more likely than the probability of one of its con-

---

[41] See Gigerenzer and Hoffrage (1995, pp. 687-688) for the algorithms (or "menus," as they call them) used and p. 693 for the result.

[42] See Cosmides and Tooby (1996).

[43] See Casscells *et al.* (1978).

[44] This example is borrowed, in part, from Hastie and Dawes (2001, p. 129).

juncts, most people feel inclined to say that it is. However, Klaus Fiedler has argued that people's tendency to commit this so-called conjunction fallacy—i.e., to ascribe the probability of a conjunct a higher probability than a conjunction in which it figures—is radically reduced if the problem is cast in frequentist terms.[45] Similarly, Gigerenzer, Hoffrage, and Klenbölting reports that our notorious and experimentally well-established tendency to over-estimate our abilities to answer factual questions correctly is diminished when we are asked to estimate the amount of questions we have answered correctly in terms of frequencies rather than probabilities.[46]

Why is this so? Cosmides and Tooby invokes an evolutionary explanation:

> In the modern world, we are awash in numerically expressed statistical information. But our hominid ancestors did not have access to the modern system of socially organized data collection, error checking, and information accumulation which has produced, for the first time in human history, reliable, numerically expressed statistical information about the world beyond individual experience. […] What *was* available in the environment in which we evolved was the encountered frequencies of actual events—for example, that we were successful 5 times out of 20 times we hunted in the north canyon. Our hominoid ancestors were immersed in a rich flow of observable frequencies that could be used to improve decision-making, given procedures that could take advantage of them. So if we have adaptations for inductive reasoning, they should take frequency information as input.[47]

So, as has been pointed out by Keith Stanovich and Richard West, there may just be "remarkably efficient mechanisms available in the

---

[45] See Fiedler (1988).

[46] See Gigerenzer, Hoffrage, and Klenbölting (1991).

[47] Cosmides and Tooby (1996, pp. 15-16).

brain—if only it was provided with the right type of representation."[48] And in the words of Gigerenzer and Hoffrage: if the evolutionary hypothesis in question is true, "Testing people's competencies for Bayesian inference with standard probability formats […] seems analogous to testing a pocket calculator's competence by feeding it binary numbers."[49]

What is important for our purposes, however, is the methodological implication that may be extracted from the experimental results—not the exact (and admittedly controversial) hypotheses about underlying psychological modules that the results may or may not warrant. As noted by Bishop and Trout, the above results suggest an obvious reasoning strategy: "When faced with a diagnosis problem framed in terms of probabilities, people should learn to represent and solve the problem in a frequency format."[50] Adapting a suggestion from Gigerenzer and Hoffrage,[51] Bishop and Trout suggest the following more specific reasoning strategy for diagnosis problems, here applied to the disease afflicting the illiterate society in the example above, and with some slight reformulations by me:

> *Diagnosis via Frequencies (DF)*
> When faced with a diagnosis problem, do as follows:
> 1. Draw up a hypothetical population of 1,000. (Literally, draw a rectangle that represents 1,000 people.)
> 2. Make a base rate cut, determining how many people (of 1,000) have the disease. (In a corner of the rectangle, color in the space representing the 10 people having the disease.)
> 3. Make a hit rate cut, determining how many of those with the disease will test positive. (In the base rate corner of the rectangle, color in the space representing the 8 true positives.)

---

[48] Stanovich and West (2002, p. 439).

[49] Gigerenzer and Hoffrage (1995, p. 699).

[50] Bishop and Trout (2005, p. 141).

[51] See Gigerenzer and Hoffrage (1995).

4. Make a false alarm cut, determining how many of those (990) without the disease will test positive. (In another corner of the rectangle, color in the space representing the 95 false positives.)
5. Determine the fraction of true positives (8) among the positives (8+95). This will tell you how many of those who test positive actually have the disease (8 in 103, or about 7.8%).[52]

As mentioned above, Bishop and Trout does not want to grant a connection—or at least not an implication—between good reasoning strategies and justified belief. However, against the background of the reconstructive framework presented in part 1 of this study, and utilized in chapter 5 through 7 to argue that the concept of justification should be reconstructed so as to be understood in terms of effective heuristics, I want to make plausible the claim that there is, in fact, such a connection. In line with the vocabulary introduced in the previous chapter, DF is a perfectly legitimate reasoning strategy. As such, its epistemic qualities have implications for the extent to which beliefs formed as a result of its employment can constructively be deemed justified or not. More specifically, remember the suggested reconstruction of justification from the previous chapter:

> *S* is justified in believing that *p* to the extent that *p* is produced by effective heuristics, i.e., heuristics that, in all or a very wide sub-set of the epistemically relevant worlds, (*i*) generate a lot of true beliefs, and (*ii*) generate a set of belief that will contain a majority of true beliefs.

So, does belief-formation resulting from DF yield justified belief? As understood here, this is an empirical question, hinging on whether or not DF satisfies the criteria cited in the reconstruction of justification. For one thing, it should be fairly uncontroversial that statistical rea-

[52] See Bishop and Trout (2005, pp. 141-142).

soning, if done correctly, has a significant tendency to yield true belief. And it is exactly in relation to the qualification "if done correctly" that our question becomes relevant, since the very truth-conductivity of statistical reasoning is what makes us interested in finding out how to reason *correctly* in statistical matters. More specifically, we are here concerned with the *comparative* question of the extent to which DF aids us in doing so to a greater extent than does its aforementioned rival, i.e., framing Bayesian reasoning in probabilistic terms. Let us refer to the probabilistic analogue of DF as DP and consider the following argument:

(8.3.3) Given accurate input, Bayes' Theorem yields accurate outputs in diagnostic matters.

This much should be beyond doubt. Indeed, it is, supposedly, by virtue of this very fact that we are at all interested in the extent to which actual diagnoses proceed along Bayesian lines and may be worried when they do not.[53] Furthermore, it should also be beyond doubt that

(8.3.4) applying Bayes' Theorem correctly (i.e., making no performance errors) has a tendency to yield a majority of true belief.

This seems to follow from the logical validity of the theorem, under the assumption that we tend to believe that which results from he methods we apply. Let us, furthermore, assume that

---

[53] Note, however, that this is certainly *not* to say that Bayesian *Epistemology* is beyond doubt. Bayesian Epistemology is a substantial normative thesis about the proper (i.e., rational) way to attain knowledge through a formal apparatus of probabilistic induction, one central component of which is the principle of conditionalization. An appreciation of the merits of employing Bayes' Theorem, however, in no way commits one to this particular normative theory.

(8.3.5) frequentist and probabilistic framings are mathematically equivalent.

This follows from the equivalence of "*x*%" and "*x* out of 100." Also, following from Gigerenzer and Hoffrage's results is the plausibility of assuming that,

(8.3.6) in the general case, a frequentist framing is less likely to give rise to performance errors than a probabilistic framing.

If so, however, we may infer that,

(8.3.7) DF has a greater tendency than DP to yield a set of belief containing a majority of true belief.

Hence, we have identified a rationale for Bishop and Trout's reasoning strategy, suggesting that, when faced with a diagnosis problem, we should learn to represent and solve the problem in a frequency format. Furthermore, if combined with the reconstruction of justification defended here, we may conclude that,

(8.3.8) other things being equal, people reasoning by way of DF are more justified in taking their resulting beliefs to be true than are people reasoning by way of DP.

Two comments are relevant. First, in analogy with our result in §8.2, a factor embedded in the "other things being equal" clause is a parity of underlying strategy determining heuristics as well as of the power of the reasoning strategy used. Consequently, (8.3.8) is perfectly compatible with claiming that people reasoning by way of DP may be more justified than people reasoning by way of DF, due to the former instantiating more effective strategy determining heuristics and/or employing more powerful reasoning strategies.

Second, (8.3.8) might seem to have a counter-intuitive ring to it, if taken to imply that the beliefs resulting from diagnostic inferences performed by subjects being so naïve of probabilistic and statistical inference that they *have* to use DF rather than DP are more justified than the inferences conducted by skilled statisticians, employing DP in accordance with what they were taught in advanced classes in graduate school. Is this implied by (8.3.8)? Yes and no. If the subject performing the inference and forming a belief as a result of it is so versed in probability theory that she rarely, if ever, makes any performance errors, she would constitute an exception to the general fact identified by Gigerenzer and Hoffrage and, hence, fall outside the scope of the "in the general case" clause in (8.3.6) above.

At the same time, however, we have also seen evidence to the effect that not even the faculty and staff at Harvard Medical School falls outside of this clause, which raises the question whether there is any *practical* point to assuming that there are very many people for which utilizing DF rather than DP would, in fact, constitute a more effective strategy. (For one thing, such an assumption seems to neglect a pretty important base rate, pertaining to the prevalence of people who fail to employ Bayes' Theorem correctly in diagnosis.) For this reason, I conclude that we may reasonably assume that, in so far as Gigerenzer and Hoffrage's results show anything at all, a person utilizing DF is generally and with few exceptions more justified in taking her beliefs to be true than is a person utilizing DP.

8.4. Elaboration, Lag, and Retention

Relevant to the idea that people under certain circumstances tend to neglect base rates is the idea, defended by Nisbett and Ross, that base rates are neglected in favor of *vivid* information. As defined by Nisbett and Ross, information is vivid to the extent that it is (*a*) emotionally interesting, (*b*) concrete and image-provoking, and (*c*) proximate in a sensory, temporal, or spatial way.[54] Admitting that the empirical evi-

[54] See Nisbett and Ross (1980, p. 45).

dence for the claim that base rates are neglected in favor of vivid information is "spotty and indirect,"[55] they support their claim by citing a study by R. M. Reyes, W. C. Thompson, and G. H. Bower on drunk driving trials, where vivid as opposed to pallid information, interestingly enough, influenced not the immediate judgment but the delayed judgment, given on the second day of the experiment.[56] More specifically, subjects exposed to vivid prosecution testimony—involving colorful descriptions of the defendant knocking over bowls of guacamole dip, "splattering guacamole all over the white shag carpet"[57]—tended to shift toward a guilty verdict, while subjects exposed to vivid defense testimony tended to shift toward a not-guilty verdict.

R. Hamil, T. D. Wilson, and Nisbett[58] have found a similar phenomenon in a study where subjects were given a description of a welfare case in an experimental setting summed up by Nisbett and Ross as follows:

> The description (condensed from an article in *the New Yorker* magazine) painted a vivid picture of social pathology. The central figure was an obese, friendly, emotional, and irresponsible Puerto Rican woman who had been on welfare for many years. Middle-aged now, she had lived with a succession of "husbands," typically also unemployed, and had borne children by each of them. Her home was a nightmare of dirty and delapidated plastic furniture bought on time at outrageous prices, filthy kitchen appliances, and cockroaches walking about in the daylight.[59]

Nisbett and Ross continues:

---

[55] Nisbett and Ross (1980, p. 47).

[56] See Reyes, Thompson, and Bower (1980).

[57] Reyes, Thompson, and Bower (1980, p. 4).

[58] See Hamil, Wilson, and Nisbett (1980).

[59] Nisbett and Ross (1980, pp. 57).

> In a second set of conditions, the article was omitted and subjects were given statistics showing that the median stay on welfare for middle-aged welfare recipients was two years and that only 10 percent of recipients remained on welfare rolls for four years or longer.[60]

As it turned out, however,

> [t]he surprising-but-pallid statistical information […] had no effect on subjects' opinions about welfare recipients. In contrast, the vivid description of one particular welfare family prompted subjects to express more unfavorable attitudes toward recipients than control subjects did. Thus highly probative but dull statistics had no effect on inferences, whereas a vivid but questionably informative case history had a substantial effect on inferences.[61]

Against the background of this and similar studies, one hypothesis delivered by Nisbett and Ross as to why vividness has such a large impact on our judgments is that vivid information is more likely to be stored and remembered than pallid information is.[62] However, Nisbett and Ross' hypothesis has been challenged on empirical grounds. In particular, and as pointed out in a review article by Shelley Taylor and Suzanne Thompson[63], several studies have directly discredited the idea that there is, in general, a strong link between retention and *concreteness*—i.e., item (*b*) in Nisbett and Ross' characterization of vividness. As for a tendency to provoke (mental) images, however, it is interesting to note that the one area in which we actually seem to get a vividness effect is in studies concerning case histories. Nevertheless,

---

[60] Nisbett and Ross (1980, pp. 57).

[61] Nisbett and Ross (1980, pp. 57-58).

[62] See Nisbett and Ross (1980, p. 45).

[63] See Taylor and Thompson (1982).

Taylor and Thompson are skeptical as to whether this effect really should be explained with recourse to vividness, especially in light of an alternative and perhaps more plausible explanation: subjects are not so much *overutilizing* vivid information as *underutilizing*—and, hence, failing to retain—base rate and other statistical information for the simple reason that they often do not understand them very well.

Perhaps this should come as no surprise in light of Gigerenzer and Hoffrage's study, given that the standard format used in psychological studies is probabilistic—not frequentist.[64] To my knowledge, there is no study that has tested this hypothesis, i.e., investigated whether the same vividness effect arises (and to the same extent) when subjects are presented with statistical information in a frequency format rather than a probabilistic format.

What *is* fairly established, however, is the more general fact that manipulations that "increase the 'depth' to which information is processed—roughly, the extent to which the subject processes the material's meaning—result in better memory," to quote Goldman.[65] Hence, T. S. Hyde and J. J. Jenkins found that subjects' abilities to remember random words were significantly better when asked to rate the pleasantness of the words as opposed to merely identify their component letters.[66] Similarly, S. A. Bobrow and G. H. Bower asked subjects to remember simple subject-verb-object sentences versus to remember sentences they themselves generated, and found that subjects had significantly better success in remembering the sentences in the latter case.[67]

By way of illustration of the possible mechanisms behind the positive influence of semantic elaboration on retention, consider the following simple but illuminating example due to John Anderson[68], where a subject is asked to remember the following sentence:

---

[64] See Gigerenzer and Hoffrage (1995).

[65] Goldman (1986, p. 212).

[66] See Hyde and Jenkins (1973).

[67] See Bobrow and Bower (1969).

[68] See Anderson (1980).

The doctor hated the lawyer.

Now, the subject attempting to remember this sentence will, most likely, store a number of other propositions in memory as well, e.g., about where the experiment took place, particular associations and feelings that it gave rise to, etc. More specifically, on deliberately *pondering* the sentence, the following propositions might arise to her:

Lawyers sue doctors for malpractice.
The lawyer had sued the doctor for malpractice.
The malpractice suit was the source of the doctor's hatred.[69]

As pointed out by Goldman, elaboration on these propositions may lead to better memory for two particular reasons. First, such elaboration may present additional *retrieval routes* for recall:

> Suppose that at the time of the test, the subject is given the word *doctor* and asked to retrieve the original sentence. The link from the node representing *doctor* to the target node, representing the sentence *The doctor hated the lawyer*, may be too weak to revive the latter. But because elaborations have led to additional associative links, recall may still be possible. From the prompt *doctor*, the subject might recall the proposition that the lawyer sued the doctor for malpractice. And from here the subject might be able to recall the target node. Thus, an alternative retrieval route may succeed if the more direct one fails.[70]

Second, elaboration may facilitate memory retrieval by way of enabling the subject to *infer* the target information. Hence, Goldman:

---

[69] See Anderson (1980, pp. 193-194)

[70] Goldman (1986, p. 213).

For example, the subject who cannot immediately recall the target sentence from the prompt *doctor* might think as follows:

> I cannot remember the target sentence but I can remember conjecturing that it was caused by the lawyer suing the doctor for malpractice and I can remember it was a sentence with a negative tone.

From this he might infer that the target sentence was *The doctor hated the lawyer*.[71]

Today, the idea that semantic elaboration has a clear, beneficial effect on retention constitutes a fairly well established hypothesis[72] and is further strengthened by recent educational research. For example, building upon research by Ference Marton and Roger Säljö on so-called "deep learning,"[73] Donald Bacon and Kim Stewart monitored the retention curve for university students enrolled in a course on consumer behavior from 8 to 101 weeks after course completion and found that information that could reasonably be assumed to have been acquired at a deeper level of understanding (for example, due to the student engaging in an elaborative process to find additional meanings and interconnections in the material) was more likely to be retained than information acquired at surface level.[74, 75]

---

[71] Goldman (1986, p. 213).

[72] See Kirchhoff, Schapiro, and Buckner (2005) for a recent study and discussion.

[73] See, e.g., Marton and Säljö (1976a, 1976b).

[74] See Bacon and Stewart (2006).

[75] Yet another relevant area of research is the relation between sleep, memory consolidation and memory strengthening. According to a popular neurological hypothesis, memory is first encoded by the hippocampus and later transferred to the cerebral cortex, particularly the neocortex, and it is believed that sleep plays a crucial role in the latter process (Stickgold, 2005). Furthermore, there are reasons to believe that there are ways to boost the mechanisms

Now, given that elaboration has a positive influence on retention, two reasoning strategies suggest themselves, one particular and one general. The general strategy may be characterized as follows:

> *Retention through elaboration—general (RE-G)*
> When faced with data the retention of which is crucial, take care to frame this data in terms that enable you to elaborate on it in depth.

Now, combine this idea with the evidence put forward by Gigerenzer and Hoffrage[76], suggesting that statistical material often is misunderstood not primarily due to humans being incapable of statistical reasoning by nature, but because probabilities—i.e., what has become the standard format of statistic material—tends to be significantly less tractable than alternative formats, such as a frequency format. This suggests a further and more specific reasoning strategy:

> *Retention through elaboration—frequentist (RE-F)*
> When faced with *statistical* data the retention of which is crucial, frame this data in terms of frequencies rather than in terms of probabilities, whenever possible.

Furthermore, as noted by Goldman, given that an increased strength in retention does not only correspond to an ability to recall a *larger* set of belief, but also an increased *reliability* in retention, the relationship between retention and depth of processing might actually suggest that

---

involved in this night-time communication between the hippocampus and the neocortex. One particularly interesting line of research here has been pursued by neuroscientist Jan Born and his research team at the University of Lübeck, who have found evidence to the effect that subjects that are presented with an odor during learning and then re-exposed to the same odor during slow-wave sleep perform significantly better on subsequent memory tasks (Born *et al.*, 2007).

[76] See Gigerenzer and Hoffrage (1995).

there are situations in which power and reliability are not conflicting desiderata.[77] Take the case of inductive inference, for example. Given that a subject's experience is fairly reliable, the more experience the subject recalls correctly, the more accurate data there is to feed into the induction. And the more accurate data there is to feed into the induction, the greater is the likelihood of a true inductive conclusion.

Returning to RE-G, let us look closer at what it means to elaborate on data in depth. Take the following paragraph:

> As conceived by classical Muslim jurists, *ijtihad* is the exertion of mental energy in the search for a legal opinion to the extent that the faculties of the jurist become incapable of further effort. In other words, *ijtihad* is the maximum effort expended by the jurist to master and apply the principles and rules of *usūl al-fiqh* (legal theory) for the purpose of discovering God's law.[78]

Now, say that you have to remember the information contained in the above paragraph. On RE-G, you should take care to frame it in terms that will allow you to elaborate on it in depth. To *not* elaborate on it in depth would involve reading through it without thinking much about its component concepts or their interrelations. To *do* elaborate on it in depth, however, would (at the very least) involve doing two things. First,

(8.4.1) identify the key concepts.

In the above passage, the key concepts are *ijtihad*, law, and God. As for the first one, one may note that the word derives from the Arabic verbal root of *jimm-ha-dal* (*jahada*, "struggle"), which is the same root as *jihad*—a shared etymology worth noting, as both words relate to a struggle or dedicated effort. As for the second concept, law, it would,

---

[77] See Goldman (1986, p. 214).

[78] Hallaq (1984, p. 3).

furthermore, be worth elaborating on the concept of a law in a Muslim context, as intimately connected to the recommendations, prohibitions, and demands of the referent of our third concept: the Muslim God or *Allah*. Second,

(8.4.2) consider the interconnections between the key concepts.

When considering the interconnections between the key concepts of the above passage, one may note that Muslim legal scholars are engaged in an explicitly non-secular pursuit, namely in a study of the word of *Allah* as dictated to his prophet Muhammad and written in the *Qur'an*; that *ijtihad* for etymological reasons may be expected to relate to a dedicated effort in this particular kind of study; and, finally, that *ijtihad*, thereby, pertains to something along the lines of a dedicated effort of a legal scholar to understand the law of God.

Now, if there is any truth to the research of Marton and Säljö, and the more recent studies by such scholars as Bacon and Stewart, a person elaborating on a piece of information along the lines of the above illustration will be more successful in subsequent memory tasks than a person that does not elaborate thus, *ceteris paribus*. To anyone who has ever employed this method, the hypothesis, undoubtedly, seems very plausible.

There are further ways in which we may increase memory power, over and above increasing depth of processing. In a study by S. A. Madigan on subjects' abilities to recall single words, forty-eight words were presented at the rate of 1.5 second, some once and some twice.[79] What Madigan found was that, in the case of the words that were presented twice, there was a robust correlation between the number of intervening words between the two presentations, or what we may refer to as *lag*, and the probability of recall on part of the subject. More specifically, there was a rapid increase in the probability of recall with increasing length of the initial lag, and the probability

[79] See Madigan (1969).

kept increasing, although less radically, with further increased lag. This is the so-called *spacing effect* and its robustness has been substantiated in more recent studies.[80] Furthermore, its obvious application should come as no surprise to anyone having first-hand experience with the difference in performance that usually comes out of repeated versus last-minute study:

> *Retention through elaboration and lag (REL)*
> When faced with data the retention of which is crucial, take care to (*a*) frame this data in terms that enable you to elaborate on them in depth, and (*b*) maximize the lag between the periods of elaboration.

Taking this reasoning strategy into account, we may add a third step to remembering the above paragraph about Muslim legal study, to the effect that we should

> (8.4.3) maximize the time elapsed between the episodes wherein you elaborate on the key concepts and their interconnections.

More than this, armed with this reasoning strategy, we may say something interesting about retention and justification. Given that the research discussed above is on the right track,

> (8.4.4) reasoning in accordance with REL increases the *power* as well as *reliability* of memory, i.e., our ability to not only remember a large set of propositions but also to remember them accurately.

---

[80] See Kahana and Howard (2005) for a recent demonstration of the beneficial effects of spacing and lag on retention. See also Bacon and Stewart (2006) for some practical suggestions as to how one may utilize these effects in educational contexts.

Note that, as here understood, remembering *accurately* implies that the proposition recalled is identical to the one that was originally believed, *not* that the proposition is true. For example, if I today come to form a belief to the effect that the Weimar Period ended in 1935 and tomorrow remember this very proposition, my memory is *accurate* (i.e., the proposition originally believed is identical to the one later recalled) even though the proposition recalled is *false* (the Weimar Period ended in 1933). However,

> (8.4.5) in so far as a majority of the propositions recalled are true (which is a question of having reliable processes feed into memory in the first place), powerful and reliable memory will significantly increase the likelihood of memory yielding a large set of belief containing a majority of true beliefs.

So, given the reconstructed concept of justification that we have been working with in the above, stating that

> *S* is justified in believing that *p* to the extent that *p* is produced by effective heuristics, i.e., heuristics that, in all or a very wide sub-set of the epistemically relevant worlds, (*i*) generate a lot of true beliefs, and (*ii*) generate a set of belief that will contain a majority of true beliefs,

we may conclude that,

> (8.4.6) given that (*a*) a majority of the propositions recalled is true and (*b*) the underlying strategy determining heuristics are effective, reasoning in accordance with REL will significantly increase the likelihood of being justified in subsequent beliefs formed by way of memory.

It is important to note that that (8.4.6), unlike (8.4.4), is *not* primarily a claim about accurate memory—i.e., of beliefs being properly preserved over time. (8.4.6) is a claim about how to form true beliefs about the world by way of memory. Note also that (8.4.6) is *not* a comparative claim, unlike (8.2.1), (8.2.2), and (8.3.8). The above research indicates that memory power and reliability is increased by elaboration and increased lag between the episodes of elaboration, and in so far as the majority of what is, thereby, remembered is also true, it is reasonable to assume that both conditions (*i*), regarding the generation of *a lot* of true belief, and (*ii*), regarding the generation of a *majority* of true beliefs, of our reconstructed concept of justification will tend to be satisfied (given effective strategy determining heuristics, of course). Naturally, this is not to deny that justification comes in degrees and that alternative strategies may yield a greater or lesser degree of justification than REL does. However, it is to say that, given that a majority of the propositions recalled is true, REL will, in many cases, yield a *sufficient* degree of justification.

This concludes our third and final demonstration of how the reconstructed concept of justification may be put to use in relation to empirical evidence about our cognitive tendencies, in order to reach hands-on advice for the epistemic inquirer to follow in her pursuit of significant truths.

8.5. Conclusion

Armed with a concept that speaks to our epistemic goals, there is no reason for epistemology to not answer to the time-honored challenge of not only describing but also guiding epistemic inquiry. The present chapter was an attempt to provide a sketch, however tentative, of how such a challenge might be answered. And although future research might very well call for a re-evaluation of this sketch, this is all as it should be. The purpose of this study has not been to provide results the basis of which are beyond all possible doubt. Rather, the aim has been to demonstrate how a certain critical study of philosophical analysis, and the methodological reconsiderations that it motivates,

may yield a framework for justification that, in turn, can be used to generate hands-on advice for epistemic inquirers reasoning on significant matters. As such, these advices are highly revisable and, indeed, *should* be revised if evidence indicates a need to do so. Hence, the claims and strategies defended are contingent upon empirical evidence that, for all we know, might turn out to be misguided tomorrow. But that is the fate of any fallible inquiry into the world we live in, and should in no way discourage the epistemologist from doing what may rightly be considered her job—to guide the epistemic inquirer in her attempts to negotiate and make sense of a complex and, in many ways, uncertain world.

Epilogue on Future Research

As noted in the beginning of chapter 8, there is a sense in which naturalistic pursuits commit themselves to a contingency in result and conditionality in formulation. So, considering the overall naturalistic approach of the present study, I find it suitable to end it by identifying a set of themes for future research that might serve to further the ambitions of the present study as well as provide the outlines of a methodology that, in the longer run, might attain the goal of guiding epistemic inquiry to an even greater extent.

The first theme is that of *epistemic value*. At the moment, there is a lively debate within epistemology—particularly as it pertains to so-called virtue epistemology—regarding the extent to which different approaches to knowledge are able to provide an explanation of why we should take knowledge to be valuable and, in particular, why we should take it to be more valuable than mere true belief.[1] The problem was, in a sense, introduced already by Plato, when he had Meno point out that there does not seem to be any *practical* difference between having knowledge and having mere true belief. For example,

[1] See Pritchard (2007) for a recent overview and Zagzebski (2004), Kvanvig (2003), and Swinburne (1999) for three influential contributions to the debate.

regardless of whether you know where Larissa is, or merely have a true belief to that effect, your chances of finding your way there will be equally great.

Apart from taking it to be an interesting problem in its own right, I also believe that providing a solution to it would enable us to get clearer on our epistemic goals and, hence, the potential multiplicity of epistemic desiderata relevant to the constructive analysis of epistemic concepts. Furthermore, it would enable us to apply the present methodological framework to KNOWLEDGE, which brings us to the second theme: *application to other epistemic concepts*. For one thing, such an inquiry would serve to evaluate the robustness of the present methodological framework for different epistemological contexts. In the present study, I have been working with an explicitly instrumentalist framework. That is, I have been assuming that the analysandum is properly understood as utilized in a context of certain norms and goals and that its normativity is to be understood in relation to these norms and goals. I suspect that there might be concepts the normativity of which does not lend itself to such a characterization. Perhaps this is the case for certain moral and aesthetic concepts such as MORALLY GOOD, RIGHT, and OBLIGATORY.

However, to the extent that we focus on instrumentalist concepts, I believe that chances are good that the methodology developed here will yield interesting and helpful results. Unlike a framework exclusively focused on concepts, a constructive analysis takes into account what may be found out about the cognitively external world—especially as revealed by what has proved to be one of our most effective ways of uncovering significant facts: science. Also, and unlike a framework that shuns concepts altogether, a constructive analysis acknowledges that analysis is done for a reason, that this reason is connected to the purpose of the concept under analysis, and that an insight into that purpose requires that we attend to that which we wish to attain by employing the concept in question.

This also connects with the third theme: *application in practical contexts*. Epistemologists tend to assume a substantial homogeneity among the specific goals of different epistemic practices. To a certain

extent, this study is no exception. While such a focus is not without its merits—for one thing, it serves to provide a unified treatment of what may reasonably be taken to be a common denominator of different epistemic pursuits—it does run the risk of ignoring the sometimes quite diverse situations and conditions of different epistemic inquires. In particular, it runs the risk of not providing a fine grained enough understanding of what we strive for in ordinary circumstances, and the particular ways in which epistemic and non-epistemic desiderata interact in naturalistic settings.

One interesting and important example here is that of evidence-based medicine. Ever since Archie Cochrane's famous series of lectures in the early 1970's, many have come to question whether diagnostic experience is sufficient ground for claims to knowledge in a medical context, or whether such experience much be accompanied by systematic evidence.[2] Although the evidence surveyed in the previous chapter regarding statistical prediction rules lends some credibility to the idea that experience should be amended thus, my point here is merely that this is exactly the kind of issue in which a constructive analysis may be helpful. Does the current concept of knowledge within medical context serve our needs well, as specified by our best theories of the typical goals and desiderata of medical practice? Is it possible to find reason for refinement or reconstruction? These questions are, clearly, of high significance, and I can at present only submit that constructive analysis presents a promising candidate to answer them.

There is a further and more general worry, however, which brings us to the fourth theme: *inter-cultural variations in reasoning patterns*. As was noted in brief above, psychologists—most notably Richard Nisbett[3]—has provided us with some reason to believe that there are substantial, cultural differences in the way people think, as in differences that can not be explained away merely with reference to different data or epistemic inputs. If true, this has important implications

---

[2] See Cochrane (1972).

[3] See Nisbett (2003).

for naturalistic epistemology. In particular, attending closely to this kind of research may enable us to better grasp either the prospects for devising more flexible reasoning strategies that have a more universal applicability, or the extent to which such strategies simply have to be customized to fit different cultural contexts.

These are but four themes for future research and I imagine that there are many more with relevance to the general ameliorative project of which the present study is a part. It is my hope, however, that I, over the course of this study, have been able to shed some light not only on the history and conditions of contemporary epistemology but also the ways in which it could be conducted in more stimulating and productive ways in the future. And only the future will tell if my particular suggestions as to how the methodology of epistemology should be improved will prove to be as constructive as I believe them to be. If I am at all on the right track, however, there are significant reasons for being optimistic about tomorrow's epistemology.

Summary in Swedish

Filosofin har under de senaste tio åren gått in i en självkritisk fas. Följaktligen har metodologiska antaganden som i årtionden – i vissa fall århundraden – tagits för självklara på sistone fått utstå omfattande kritik. Denna studie utgör ett inlägg i denna kritiska granskning av filosofisk metodologi, speciellt med avseende på kunskapsteoretisk metodologi. Den börjar sin undersökning i filosofins vagga med Platons sokratiska metod, för att sedan blottlägga en omfattande metodologisk likhet mellan denna och modern filosofi, som består i att filosofiska teorier tar formen av *definitioner* som väsentligen försvaras eller förkastas med hänvisning till *intuitioner*.

Det föreslås att denna metodologiska kontinuitet ska förstås i termer av en strukturell likhet mellan, å ena sidan, Platons teori om Former och rationell insikt och, å andra sidan, en modern men implicit teori om begrepp – den moderna filosofins huvudsakliga undersökningsobjekt. Enligt denna teori representeras begrepp bäst via enkla och skarpt skurna, nödvändiga och tillräckliga villkor, tillgängliga för filosofen genom ett (ofta introspektivt) studium av våra kategoriseringstendenser. Det första problemet med denna teori är att modern psykologisk forskning inte ger oss några skäl att anta att begrepp ska förstås på detta sätt, varför en metodologisk förbättring i termer

av så kallad prototypteori föreslås. Att på detta sätt inkorporera psykologisk forskning väcker den vidare frågan huruvida en traditionell länsstolsmetod verkligen erbjuder en lämpligare metodologisk utgångspunkt än modern psykologisk forskning. Denna fråga blir dock relevant först under antagandet att filosofer bör söka fullständigt *uttömmande* analyser av våra epistemiska begrepp. Jag argumenterar för att detta antagande är ogrundat och att en förståelse av våra begrepp främst är intressant som en approximativ utgångspunkt i sökandet efter begrepp som underlättar våra kognitiva åtaganden *till en större grad* än vad våra nuvarande begrepp gör.

Mer specifikt föreslås det att kunskapsteoretisk metodologi bör förstås i termer av två uppgifter, indelade i två deluppgifter vardera. Den första uppgiften är *deskriptiv* och går ut på att (*a*) *identifiera* undersökningsobjektet genom att plocka ut ett antal paradigmatiska exempel, för att sedan (*b*) *aggregera* dessa objekts egenskaper med syftet att finna en rättvis karaktärisering av det epistemiska fenomenet i fråga. Utöver att aggregeringen är väsentligen empirisk bör det också noteras att identifikationen knappast förutsätter en särskilt omfattande begreppsanalys. Jag föreslår istället att den senare ska förstås i termer av *meningsanalys*, som istället för att arbeta med uttömmande definitioner arbetar med så kallade *stereotyper* – listor av typiska egenskaper som varken bestämmer referensen eller nödvändigtvis utgör en korrekt teori för denna, men som ändå kan sägas fånga en viktig, kognitiv aspekt av det korresponderande begreppet.

Varken vikten av empiri eller den uppenbart modesta meningsanalysen innebär dock att begrepp i sig själv inte spelar någon central roll inom kunskapsteoretisk analys, vilket leder oss till filosofins andra och *normativa* uppgift: att (*a*) *utvärdera* i vilken mån våra begrepp verkligen underlättar våra kognitiva åtaganden samt (*b*) *förbättra* dem i den mån de kan tänkas göra detta till en större grad än vad de faktiskt gör. Inte heller dessa moment förutsätter dock traditionell begreppsanalys, då denna tenderar att ignorera teleologiska frågor rörande *varför* vi kategoriserar världen som vi gör och för vilka syften vi därmed använder våra begrepp, genom att uteslutande fokusera på applikationsvillkor och *att* vi kategoriserar världen på ett visst

sätt. Jag förslår därför en mer rimlig typ av analys som jag döper till *begrepplig syftesanalys*. Denna typ av analys syftar till att blottlägga den roll som våra begrepp kan förväntas spela inom våra diverse kognitiva åtaganden, för att därmed föregå samt vägleda en närmare undersökning av de sätt på vilket våra begrepp skulle kunna förbättras och därigenom öka våra chanser att nå våra mål.

Jag väljer att kalla denna metodologi för *konstruktiv analys*. Namnet är valt mot bakgrund av att analysen i fråga lämnar rum inte bara för mindre förfinanden utan även för *rekonstruerandet* av begrepp i den mån de inte fyller sin funktion, samt för att analysen därigenom även kan förväntas spela en *konstruktiv roll* i relation till praktisk, epistemisk verksamhet. Detta illustreras i studiens andra del där den konstruktiva analysen tillämpas på ett fenomen som uppmärksammats mycket inom modern kunskapsteorisk diskussion: epistemiskt berättigande.

En närmare betraktelse över två inflytelserika teorier avslöjar här flera problem. Det första förslaget – tanken att berättigande består i att ha uppfyllt vissa epistemiska plikter – faller på faktumet att vi inte har någon makt över formandet av våra trosföreställningar och följaktligen inte heller kan klandras för att vi formar vissa trosföreställningar snarare än andra. Det andra förslaget kan förstås som en mindre rekonstruktion av det förra, genom att den (åtminstone i vissa fall) håller fast vid talet om epistemiska plikter men hävdar att objektet för epistemisk utvärdering inte är formandet av trosföreställningar utan snarare introspektionsakter. Berättigande består utifrån detta förslag nämligen i att ha skärskådat grunden för sina trosföreställningar genom introspektion. Även detta förslag stöter dock på allvarliga problem: enligt modern kogntivpsykologisk forskning har vi sällan introspektiv tillgång till varför vi tror vad vi tror. Dessutom har vi, i många fall där vi faktiskt har sådan tillgång, en tendens att ge en allt annat än korrekt utvärdering av giltigheten hos våra skäl.

Detta motiverar utarbetandet av en mer rimlig rekonstruktion som inte lider av ovanstående problem och som därigenom kan utgöra ett bättre verktyg givet våra specifika epistemiska förutsättningar och desiderata. Det poängteras än en gång att empirisk forskning är

högst relevant för förståelsen av dessa förutsättningar och desiderata och i en närmare undersökning karaktäriserar jag de senare i termer av ett sökande efter *signifikanta sanningar*, dvs. sanningar som anknyter till vad vi anser vara viktiga frågor, problem, hypoteser o. dyl. Jag utarbetar sedan en rekonstruktion som väsentligen består i att identifiera berättigande med användandet av *effektiva tankemönster*, förstådda som antingen medvetna resonemangsstrategier eller omedvetna tankeprocesser som tenderar att generera uppsättningar av trosföreställningar som innehåller (*a*) *många* sanna trosföreställningar samt (*b*) en *majoritet* av sanna trosföreställningar.

Detta leder oss till avhandlingens sista kapitel där jag fokuserar på våra epistemiska förutsättningar och specifikt på möjligheten att använda vårt rekonstruerade begrepp för att utveckla förbättrade resonemangsstrategier. Mot bakgrund av en studie av relevant psykologisk litteratur, tillämpar jag detta begrepp inom tre specifika områden: prediktion, diagnostik och retention, i sin mest generella form svarandes mot frågorna *Vad kommer att vara fallet? Vad är fallet?* och *Vad var fallet?* Mer specifikt försöker jag illustrera hur den föreslagna rekonstruktionen kan användas för att, i kombination med empirisk forskning, generera direkta rekommendationer som syftar till att underlätta våra epistemiska strävanden. Till exempel diskuteras forskning som indikerar att även en specifik men mycket kort statistisk träning minimerar riskerna för en rad ökända statistiska felslut och därmed underlättar konstruerandet av bevisligen högst effektiva prediktionsregler; hur omformulerandet av diagnostiska problem i termer av frekvenser snarare än sannolikheter radikalt minskar riskerna för diagnostiska felslut; samt specifika strategier för att förbättra vår förmåga att minnas även högst detaljerad information.

Givet en förväntad filosofisk skepsis rörande den mån i vilket denna typ av resultat alls har några implikationer för vårt berättigandebegrepp argumenterar jag som följer: Givet (*a*) de ovan berörda problemen med våra filosofiska teorier om berättigande, speciellt med avseende på hur de lämpar sig dåligt vad gäller att underlätta vår strävan efter (*b*) de specifika mål vi önskar uppnå som epistemiska varelser, samt (*c*) faktumet att ett begrepps *raison d'âitre* är (eller åtminstone

bör vara) direkt proportionellt till den mån i vilket det fyller sin funktion, har vi goda anledningar att i vissa sammanhang helt enkelt låta det rehabiliterade berättigandebegreppet ersätta det (eller de) begrepp vi råkar ha ärvt från våra kunskapsteoretiska föregångare.

Därmed tar jag mig även för att ha visat hur kunskapsteorin, förstådd i termer av konstruktiv analys, kan leva upp till den traditionella och minst sagt hedervärda utmaningen att inte bara beskriva utan även vägleda faktiskt kunskapssökande verksamhet—en utmaning som bör ligga varje naturalistisk kunskapsteori nära hjärtat.

Bibliography

Ahlström, K., (2006), "Realism and Scientific Failure," in *Essays Dedicated to Dag Westerståhl on his 60th Birthday*, Göteborg.

Alston, W. P., (1985), "Concepts of Epistemic Justification," *The Monist* 68, pp. 57-89.

——— (1986), "Internalism and Externalism," *Philosophical Topics* 14, pp. 179–221.

——— (1989), *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.

——— (1993), "Epistemic Desiderata," *Philosophy and Phenomenological Research* LIII (3), pp. 527-551.

——— (2005), *Beyond "Justification": Dimensions of Epistemic Evaluation*, Ithaca, NY: Cornell University Press.

Anderson, J., (1980), *Cognitive Psychology and its Implications*, San Francisco, CA: W. H. Freeman.

Armor, D. A., (1998), *The Illusion of Objectivity: A Bias in the Perception of Freedom from Bias*, doctoral dissertation, University of California, Los Angeles, CA.

Armstrong, D., (1973), *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

Ashenfelter, O., Ashmore, D., and Lalonde, R., (1995), "Bordeaux Wine Vintage Quality and the Weather," *Chance* 8 (4), pp. 7-14.

Audi, R., (1998), Epistemology: *A Contemporary Introduction to the Theory of Knowledge*, New York: Routledge.

Ayer, A. J., (1956), *The Problem of Knowledge*, London: Macmillan.

Bacon, D. R., and Stewart, K. A., (2006), "How Fast do Students Forget What They Learn in Consumer Behavior? A Longitudinal Study," *Journal of Marketing Education* 28, pp. 181-192.

Bealer, G., (1998), "Intuition and the Autonomy of Philosophy," in M. R. DePaul and W. Ramsey (1998), pp. 201-240.

Beebe, J. R., (2004), "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis," *Noûs* 38 (1), pp. 177-195.

Bernecker, S., (2006), "Prospects for Epistemic Compatibilism," *Philosophical Studies*, 130, pp. 81-104.

Bishop, M., and Trout., J. D., (2005), *Epistemology and the Psychology of Human Judgment*, Oxford: Oxford University Press.

Bobrow, D. G., and Bower, G. H., (1969), "Comprehension and Recall of Sentences," *Journal of Experimental Psychology* 80, pp. 455-461.

Boghossian, P., (1996), "Analyticity Reconsidered," *Noûs* 30 (3), pp. 360-391.

BonJour, L., (1985), *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

——— (1998), *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*, Cambridge: Cambridge University Press.

BonJour, L., and Sosa, E., (2003), *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, Malden, MA: Blackwell Publishing.

Borges, J. L., (1999), *Borges: Collected Fiction*, (trans. by A. Hurley), New York, NY: Penguin Books.

Born, J., Rasch, B., Büchel, C., and Gais, S., (2007) "Odor Cues During Slow-Wave Sleep Prompt Declarative Memory Consolidation." *Science* 315, pp. 1426-1429.

Boyd, R., (1984), "The Current Status of Scientific Realism," In J. Leplin, ed., *Scientific Realism*. Berkeley & Los Angeles, CA: University of California Press, pp. 41-82.

Burge, T., (1986), "Individualism and Psychology," *Philosophical Review* 95, pp. 3-45.

Carey, S., (1999), "Knowledge Acquisition: Enrichment or Conceptual Change?" in E. Margolis and S. Laurence (1999), pp. 459-487. Originally published in S. Carey and R. Geldman, eds., *The Epigenesis of Mind: Essays on Biology and Cognition*, Hillside, NJ: Lawrence Erlbaum Associates, Inc., 1991.

Carnap, R., (1950), *Logical Foundations of Probability*, Chicago, IL: Chicago University Press.

Carpenter, R. G., and Emory, J. L., (1977), "Final Results of Study of Infants at Risk of Sudden Infant Death," *Nature* 268, 724-725.

Carroll, J. S., Wiener, R. L., Coates, D., Galegher, J., and Alibrio, J. J., (1982) "Evaluation, Diagnosis, and Prediction in Parole Decision Making," *Law & Society Review* 17 (1), pp. 199-228

Casscells, W., Schoenberger, A., and Grayboys, T., (1978), "Interpretation by physicians of Clinical Laboratory Results," *New England Journal of Medicine* 299, pp. 999-1001.

Chapman, G. B., and Bornstein, B. H., (1996), "The More You Ask For, The More You Get: Anchoring in Personal Injury Verdicts," *Applied Cognitive Psychology* 10, pp. 519-540.

Chisholm, R. M., (1957), *Perceiving: A Philosophical Study*, Ithaca, New York: Cornell University Press.

——— (1968), "Lewis' Ethics of Belief," in P. Shilpp, ed., *The Philosophy of C. I. Lewis*, LaSalle: IL: Open Court.

——— (1977), *Theory of Knowledge*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall.

Cochrane A. L., (1972), *Effectiveness and Efficiency: Random Reflections on Health services*, London: Nuffield Provincial Hospitals Trust.

Conee, E., (1988), "The Basic Nature of Epistemic Justification," *The Monist* 71.

Conee, E., and Feldman, R., (1998), "The Generality Problem for Reliabilism," *Philosophical Studies* 89, pp. 1-29.

Cornman, J. W., Lehrer, K., and Pappas, G., (1982), *Philosophical Problems and Arguments: An Introduction*, 3rd ed., New York: Macmillan.

Cosmides, L., and Tooby, J., (1996), "Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty," *Cognition* 58, pp. 1-73.

Craig, E., (1990), *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*, Oxford: Clarendon Press.

Damasio, A., (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, NY: Philosophical Library.

Daniels, N., (1979), "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *The Journal of Philosophy* 76 (5), pp. 256-282.

David, M., (2005), "Truth as the Primary Epistemic Goal: A Working Hypothesis," in M. Steup and E. Sosa, eds., *Contemporary Debates in Epistemology*, Malden, MA: Blackwell Publishing.

Davidson, D., (1980), "Actions, Reasons, and Causes," in his *Essays on Actions and Events*, Oxford: Clarendon Press.

Dawes, R. M., and Corrigan, B., (1974), "Linear Models in Decision Making," *Psychological Bulletin* 81, pp. 95-106.

Dawes, R., Faust, D., and Meehl, P., (2002), "Clinical versus Actuarial Judgment," in T. Gilovich, D. Griffin, and D. Kahneman (2002), pp. 716-729.

DePaul, M. R., (1998), "Why Bother With Reflective Equilibrium?" in M. R. DePaul and W. Ramsey (1998), pp. 293-309.

DePaul, M. R., and Ramsey, W., eds., (1998), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman & Littlefield Publishers, Inc.

Descartes, R., (1955), "Meditations of First Philosophy," in E. S. Haldane and G. R. T. Ross, eds., *Philosophical Works of Descartes*, vol. 1, Dover Publications.

——— (1988a), "Rules for the Direction of Our Native Intelligence," in J. Cottingham, R. Stoothoff, and D. Murdoc, eds., *Descartes: Selected Philosophical Writings*, Cambridge: Cambridge University Press, pp. 1-19.

——— (1988b), "Meditations on First Philosophy," in J. Cottingham, R. Stoothoff, and D. Murdoc, eds., *Descartes: Selected Philosophical Writings*, Cambridge: Cambridge University Press, pp. 73-122.

Devitt, M., (2006), "Intuitions in Linguistics," *The British Journal for the Philosophy of Science* 57 (3), pp. 481-513.

Diener, E., (2000), "Subjective Well-Being: The Science of Happiness, and a Proposal for a National Index," *American Psychologist* 55, pp. 34-43.

Diener, E., and Oishi, S., (2000), "Money and Happiness: Income and Subjective Well-Being Across Nations," in E. Diener and E. M. Suh, eds., *Culture and Subjective Well-Being*, Cambridge, MA: The MIT Press, pp. 185-218.

Diener, E., and Seligman, M. E. P., (2002), "Very Happy People," *Psychological Science* 13, pp. 80-83.

Dretske, F., (1988), *Explaining Behavior*, Cambridge, MA: The MIT Press.

Dummett, M., (1991), *The Logical Basis of Metaphysics*, Cambridge: Harvard University Press.

Eddy, D. M., (1982), "Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities," in D. Kahneman, P. Slovic, and A. Tversky (1982), pp. 249-267.

Einhorn, H. J., and Hogarth, R. M., (1975) "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Decision Processes* 12, pp. 171-192.

Elga, A., (2005), "On Overrating Oneself… and Knowing It," *Philosophical Studies* 123, pp. 115-124.

Faust, D., and Ziskin, J., (1988), "The Expert Witness in Psychology and Psychiatry," *Science* 241, pp. 31-35.

Feldman, R., (1985), "Reliability and Justification," *The Monist* 68, pp. 159-174.

——— (1993): "Proper Functionalism," *Noûs* 27, no. 1, pp. 34–50.

——— (2000), "The Ethics of Belief," *Philosophy and Phenomenological Research* 60 (3), pp. 667-695.

——— (2001), "Voluntary Belief and Epistemic Evaluation," in M. Steup, ed., *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, Oxford: Oxford University Press, pp. 77-92.

Fiedler, K., (1988), "The Dependence of the Conjunction Fallacy on Subtle Linguistic Factors," *Psychological Research* 50, pp. 123-129.

Field, H., (1973), "Theory Change and the Indeterminacy of Reference," *The Journal of Philosophy* 70 (14), pp. 462-481.

Fine, A., (1986), "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science," *Mind* 95, pp. 149-179.

Fischhoff, B., (2002), "Heuristics and Biases in Application," in T. Gilovich, T. Griffin, and D. Kahneman (2002), pp. 730-748.

Fodor, J., (2003), *Hume Variations*, Oxford: Oxford University Press.

——— (1981), "The Present Status of the Innateness Controversy," in his *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge: MA: The MIT Press.

Foley, R., (1987), *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.

Fong, G. T., Krantz, D. H., and Nisbett, R. E., (manuscript), "Improving Inductive Reasoning Through Statistical Training," unpublished manuscript, University of Michigan, Ann Arbor, MI.

Frankfurt, H. G., (1969), "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66 (23), pp. 829-839.

Frege, G., (1953), *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, (trans. by J. L. Austin), Oxford: Blackwell.

Frey, B. S., and Stutzer, A., (2002), *Happiness and Economics*, Princeton, NJ: Princeton University Press.

Gettier, E., (1963), "Is Justified True Belief Knowledge?" *Analysis* 23, pp. 121-123.

Gibbons, J., (2006), "Access Externalism," *Mind*, vol. 115, no. 457, pp. 19-39.

Giere, R. N., (1999), *Science Without Laws*, Chicago, IL: University of Chicago Press.

Gigerenzer, G., (1996), "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)," *Psychological Review* 103, pp. 592-596.

Gigerenzer, G., and Hoffrage, U., (1995), "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats," *Psychological Review* 102 (4), pp. 684-704.

Gigerenzer, G., and Selten, R., eds., (2001), *Bounded Rationality: The Adaptive Toolbox*, Cambridge, MA: The MIT Press.

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H., (1991), "Probabilistic Mental Models: A Brunswikian Theory of Confidence," *Psychological Review* 98, pp. 506-528.

Gigerenzer, G., Todd, P. M., and the ABC Research Group, eds., (1999), *Simple Heuristics that Make Us Smart*, Oxford: Oxford University Press.

Gilbert, D. T., Pinel., E. C., Wilson, T. D., Blumberg, S. J., and Wheatley, T. P., (2002), "Durability Bias in Affective Forecasting," in T. Gilovich, D. Griffin, and D., Kahneman (2002), pp. 292-312.

Gilovich, T., Griffin, D., and Kahneman, D., (2002), eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.

Goldberg, L. R., (1970), "Man Versus Model of Man: A Rationale, Plus Some Evidence, For A Method Of Improving On Clinical Inferences," *Psychological Bulletin* 73, pp. 422-432.

Goldman, A., (1986), *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.

——— (1992a), *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: The MIT Press.

——— (1992b), "Epistemic Folkways and Scientific Epistemology," in Goldman (1992a), pp. 155-175.

——— (1999), "Internalism Exposed," *Journal of Philosophy* 96, pp. 271-293.

——— (2001), "Replies to the Contributors," *Philosophical Topics* 29, pp. 461-511.

——— (2007), "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status," *Grazer Philosophische Studien* 74, pp. 1-26.

Goldman, A., and Pust, J., (1998), "Philosophical Theory and Intuitional Evidence," in M. R. DePaul and W. Ramsey (1998), pp. 179-197.

Goodman, N., (1983), *Fact, Fiction, and Forecast*, 4th ed., Cambridge, MA: Harvard University Press. 1st ed. originally published in 1955.

Griffiths, P. E., (1997), *What Emotions Really Are: The Problem of Psychological Categories*, Chicago and London: The University of Chicago Press.

Haack, S., (1976), "The Pragmatist Theory of Truth," *British Journal for the Philosophy of Science* 27, pp. 231-249.

Hallaq, W. B., (1984), "Was the Gate of *Ijtihad* Closed?" *International Journal of Middle East Studies* 16 (1), pp. 3-41.

Hamil, R., Wilson, T. D., and Nisbett, R. E., (1980), "Insensitivity to Sample Bias: Generalizing form Atypical Instances," *Journal of Personality and Social Psychology* 39, pp. 578-589.

Hastie, R., and Dawes, R. M., (2001), *Rational Choice in an Uncertain World*, Thousand Oaks, CA: Sage Publications.

Hastie, R., Schkade, D. A., and Payne, J. W., (1999), "Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards," *Law and Human Behavior* 23, pp. 445-470.

Henderson, D., and Horgan, T. E., (2001), "Practicing Safe Epistemology," *Philosophical Studies* 102, pp. 227-258.

Howard, J. W., and Dawes, R. M., (1976), "Linear Prediction of Marital Happiness," *Personality and Social Psychology Bulletin* 2 (4), pp. 478-480.

Hyde, T. A., and Jenkins, J. J., (1973), "Recall for Words as a Function of Semantic, Graphic, and Syntactic Orienting Tasks," *Journal of Verbal Learning and Verbal Behavior* 12 (5), pp. 471-480.

Jackson, F., (1998), *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon Press.

Kahana, M. J., and Howard, M. W., (2005), "Spacing and Lag Effects in Free Recall of Pure Lists," *Psychonomic Bulletin & Review* 12 (1), pp. 159-164.

Kahneman, D., and Tversky, A., (1972), "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology* 3, pp. 430-454.

——— (1973), " On the Psychology of Prediction," *Psychological Review* 80, pp. 237-251.

Kahneman, D., Slovic, P., and Tversky, A., eds., (1982), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge, MA: Cambridge University Press.

Kaplan, M., (1991), "Epistemology on Holiday," *The Journal of Philosophy* 88 (3), pp. 132-154.

Kappel, K., (manuscript), "A Diagnosis and Resolution to the Generality Problem," unpublished manuscript.

Keil, F. C., (1989), *Concepts, Kinds and Cognitive Development*, Cambridge, MA: the MIT Press.

Kirchhoff, B. A., Schapiro, M. L., and Buckner, R. L., (2005), "Orthographic Distinctiveness and Semantic Elaboration Provide Separate Contributions to Memory," *Journal of Cognitive Neuroscience* 17 (12), pp. 1841-1854.

Kitcher, P., (1992), "The Naturalists Return," *The Philosophical Review* 101 (1), pp. 53-114.

——— (2001), *Science, Truth, and Democracy*, Oxford: Oxford University Press.

Kornblith, H., (1993), *Inductive Inference and Its Natural Ground*, The MIT Press.

——— (2001), "Epistemic Obligation and the Possibility of Internalism," in A. Fairweather and L. Zagzebski, eds., *Virtue Epistemology: Essays on*

Epistemic Virtue and Responsibility, Oxford: Oxford University Press, pp. 231-248.

——— (2002), *Knowledge and Its Place in Nature*, Oxford: Oxford University Press.

——— (2006), "Appeals to Intuition and the Ambitions of Epistemology," in S. Hetherington, ed., *Epistemology Futures*, Oxford: Oxford University Press, 2006, pp. 10-25.

——— (2007), "Naturalism and Intuitions," *Grazer Philosophische Studien* 74, pp. 27-49.

——— (forthcoming a), "How to Refer to Artifacts," in E. Margolis and S. Laurence, eds., *Creations of the Mind: Essays on Artifacts and their Representation*, Oxford: Oxford University Press, forthcoming.

——— (forthcoming b), "What Reflective Endorsement Cannot Do," forthcoming in *Philosophy and Phenomenological Research*.

Krantz, D. H., Fong, G. T., and Nisbett, R. E., (manuscript), "Formal Training Improves the Application of Statistical Heuristics to Everyday Problems," unpublished manuscript, Murray Hill, NJ: Bell Laboratories.

Kripke, S., (1980), *Naming and Necessity*, Cambridge, MA: Harvard University Press. Originally published in D. Davidson and G. Harman, eds., *Semantics of Natural Language*, Dordrecht: Reidel, 1972, pp. 253-355.

Kvanvig, J., (2003) *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.

Langford, C. (1942), "The Notion of Analysis in Moore," in P. Schilpp, ed., *The Philosophy of G.E. Moore*, Chicago: Open Court.

LaPorte, Joseph (2004), *Natural Kinds and Conceptual Change*, Cambridge: Cambridge University Press.

Laurence, S., and Margolis, E., (2003), "Concepts and Conceptual Analysis," *Philosophy and Phenomenological Research* LXVII (2), pp. 253-282.

——— eds., (1999), *Concepts: Core Readings*, Cambridge, MA: The MIT Press.

Lehrer, K., and Cohen, S., (1983), "Justification, Truth, and Coherence," *Synthese* 55, pp. 191-207.

Lewis, D., (1973), *Counterfactuals*. Oxford: Blackwell.

Locke, J., (1996), *An Essay Concerning Human Understanding*, Indianapolis, IN: Hackett Publishing Company.

Lord, C., Ross, L., and Lepper, M., (1979), "Biased Assimilation and Attitude Polarization: The Effects of Priori Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology* 34.

Lovie, A. D., and Lovie, P., (1986), "The Flat Maximum Effect and Linear Scoring Models for Prediction," *Journal of Forecasting* 5 (3), pp. 159-168.

Lycan, W., (2006), "On the Gettier Problem Problem," in S. Hetherington, ed., *Epistemology Futures*, Oxford University Press, 2006, pp. 148-168.

Madigan, S. A., (1969), "Intraserial Repetition and Coding Processes in Free Recall," *Journal of Verbal Learning and Verbal Behavior* 8, pp. 828-835.

Mason, S. F., (1962), *A History of the Sciences*, New York, NY: Collier Books.

Marton, F., and Säljö, R., (1976a), "On Qualitative Differences in Learning: 1. Outcome and Process," *British Journal of Educational Psychology* 46, pp. 4-11.

——— (1976b), "On Qualitative Differences in Learning: 2. Outcome as a Function of Learners Conception of Task," *British Journal of Educational Psychology* 46, pp. 115-127.

McDowell, J., (1973), *Plato's Theaetetus*, Oxford: Clarendon Press.

Meehl, P., (1954), *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis, MN: University of Minneapolis Press.

——— (1986), "Causes and Effects of My Disturbing Little Book," *Journal of Personality Assessment* 50, pp. 370-375.

Moser, P., (1985), *Empirical Justification*, Dordrecht: Kluwer Publishing Company.

Murphy, G. L., (1993), "A Rational theory of Concepts," *Psychology of Learning and Motivation* 29, pp. 327-359.

——— (2002), *The Big Book of Concepts*, Cambridge, MA: The MIT Press.

Murphy, G. L., and Medin, D., (1985), "The Role of Theories in Conceptual Coherence," *Psychological Review* 92 (3), pp. 289-316.

Naess, A., (1938), "'Truth' as it is Conceived by Those Who are not Philosophers," Oslo.

——— (1953), *Interpretation and Preciseness*, Oslo.

Nisbett, R. E., (2003), *The Geography of Thought: How Asians and Westerners Think Differently… and Why*, New York, NY: Free Press.

Nisbett, R. E., and Borgida, E., (1975), "Attribution and the Social Psychology of Prediction," *Journal of Personality and Social Psychology* 32, pp. 932-943.

Nisbett, R. E., and Ross, L., (1980), *Human Inference*: *Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R. E., and Wilson, T., (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84, pp. 231-259.

Nisbett, R. E., Krantz, D. H., Jepson, C., and Kunda, Z., (2002), "The Use of Statistical Heuristics in Everyday Inductive Reasoning," in T. Gilovich, D. Griffin, and D. Kahneman (2002), pp. 510-533.

Nozick, R., (1981), *Philosophical Explanations*, Cambridge, MA: Harvard University Press.

Parfit, D., (1984). *Reasons and Persons*, Oxford: Clarendon Press.

Pinker, S., (2007), *The Stuff of Thought*, New York, NY: Viking.

——— (1994), *The Language Instinct: The New Science of Language and Mind*, London: Penguin.

Plato, (1953), *Dialogues*, 4th ed., (trans. by B. Jowett), Oxford: Clarendon Press.

Plato, (1970), *Protagoras and Meno* (trans. by W. K. C. Guthrie), Baltimore, MD: Penguin Books.

Plantinga, A., (1988), "Positive Epistemic Status and Proper Functioning," *Philosophical Perspectives* 2, pp. 1-50.

——— (1990), "Justification in the 20th Century," *Philosophy and Phenomenological Research* 50 (supplement), pp. 45-71.

Pritchard, D., (2007), "Recent Work on Epistemic Value," *American Philosophical Quarterly* 44 (2), pp. 85-110.

Pronin, E., Lin, D. Y., and Ross, L., (2002), "The Bias Blind Spot: Perceptions of Bias in Self versus Others, *Personality and Social Psychology Bulletin,* 28, pp. 369-381.

Pronin, E., Puccio, C., and Ross, L., (2002), "Understanding Misunderstanding: Social Psychological Perspectives," in T. Gilovich, D. Griffin, and D. Kahneman (2002), pp. 636-664.

Pryor, J., (2000), "The Skeptic and the Dogmatist," *Noûs* 34 (4), pp. 517-549.

Puccio, C, and Ross., L., (manuscript), "Real versus Perceived Ideological Differences: Can We Close the Gap?" unpublished manuscript, Stanford University, Stanford, CA.

Putnam, H., (1975a), "Is Semantics Possible?" in *Mind, Language, and Reality: Philosophical Papers, vol. 2*, Cambridge: Cambridge University Press, pp. 139-152.

——— (1975b), "The Meaning of 'Meaning'," in *Mind, Language, and Reality: Philosophical Papers vol. 2*, Cambridge: Cambridge University Press, pp. 215-271.

——— (1975c), "The Analytic and the Synthetic," in *Mind, Language, and Reality: Philosophical Papers vol. 2*, Cambridge: Cambridge University Press, pp. 33-69.

Quattrone, G. A., Lawrence, C. P., Warren, D. L., Souza-Silva, K., Finkel, S. E., and Andrus, D. E., (manuscript), *Explorations in Anchoring: The Effects of Prior Range, Anchor Extremity, and Suggestive Hints*, unpublished manuscript, Stanford University, Stanford, CA.

Quine, W. V. O., (1951), "Two Dogmas of Empiricism," *The Philosophical Review* 60, pp. 20-43.

——— (1960), *Word and Object*, Cambridge, MA: The MIT Press.

——— (1969), *Ontological Relativity and Other Essays*, New York, NY: Columbia University Press.

Ramsey, F. P., (1931), *The Foundations of Mathematics, and Other Logical Essays*, London: Routledge and Kegan Paul.

Ramsey, W., (1998), "Prototypes and Conceptual Analysis," in M. R. DePaul & W. Ramsey (1998), pp. 161-178.

Rawls, J., (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.

——— (1974/5), "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association*, XLVII, pp. 5-28.

Reyes, R. M., Thompson, W. C., and Bower, G. H., (1980), "Judgmental Biases Resulting from Differing Availabilities of Arguments," *Journal of Personality and Social Psychology* 37, pp. 2-12.

Rips, L. J., Shoben, F. J., and Smith, F. E., (1973), "Semantic Distance and the Verification of Semantic Relations," *Journal of Verbal Learning and Verbal Behavior* 12, pp. 1-20.

Rosch, E., (1977), "Human Categorization," in N. Warren, ed., *Studies in Cross-Cultural Psychology*, vol. 1, London: Academic Press.

Rosch, E., and Mervis, C., (1975), "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology* 8, pp. 382-439.

Samuels, R., Stich, S., and Faucher, L., (2004), "Reason and Rationality," in I. Niiniluoto, M. Sintonen., and J. Wolenski, eds., *Handbook of Epistemology*, Dordrecht: Kluwer Academic Press, pp. 131-181.

Samuels, R., Stich, S., and Tremoulet, P. D., (1999), "Rethinking Rationality: From Bleak Implications to Darwinian Modules," in E. Lepore and Z. Pylyschyn, eds., *Rutgers University Invitation to Cognitive Science*, Oxford: Basil Blackwell.

Sawyer, J., (1966), "Measurement and Prediction, Clinical and Statistical," *Psychological Bulletin* 1, pp. 54-87.

Schmitt, F., (1992), *Knowledge and Belief*, London: Routledge and Kegan Paul.

Sebo, J., (2006), "Is a Real Ethics of Belief Possible?" unpublished manuscript presented at *the 4th Biennial Rochester Epistemology Conference* in Rochester, NY, October 2006.

Segal, G. M. A., (2000), *A Slim Book About Narrow Content*, Cambridge, MA: The MIT Press.

Sen, A., (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation*, Oxford: Oxford University Press.

Shope, R. K., (1983), *The Analysis of Knowing: A Decade of Research*, Princeton: Princeton University Press.

Smart, J. J. C., (1968), *Between Science and Philosophy*, New York, NY: Random House.

Smith, E., and Medin, D., (1981), *Concepts and Categories*, Cambridge, MA: The MIT Press.

Sosa, E., (2003), "The Place of Truth in Epistemology," in M. DePaul and L. Zagzebski, *Intellectual Virtue: Perspectives From Ethics and Epistemology*, Oxford: Oxford University Press, pp. 155-179.

Stanford, P. K., and Kitcher, P., (2000), "Refining the Causal Theory of Reference for Natural Kind Terms," *Philosophical Studies* 97, pp. 99-129.

Stanovich, K. E., and West, R. F., (2002), "Individual Differences in Reasoning: Implications for the Rationality Debate," in T. Gilovich, D. Griffin, and D. Kahneman (2002), pp. 421-439.

Stich, S., (1998), "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity," in M. R. DePaul and W. Ramsey (1998), pp. 95-112.

Stickgold, R., (2005) "Sleep-Dependent Memory Consolidation," *Nature* 437, pp. 1272-1278.

Stillwell, W., Barron, F., and Edwards, W., (1983), "Evaluating Credit Applications: A Validation of Multiattribute Utility Weight Elicitation Techniques," *Organizational Behavior and Human Performance* 32, pp. 87−108.

Strawson, P. F., and Grice, H. P., (1956), "In Defense of a Dogma," *Philosophical Review* 65, pp. 141-158.

Swain, S., Alexander, J., and Weinberg, J. M., (manuscript), "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp," unpublished manuscript, available at http://www.indiana.edu/~eel/.

Swinburne, R., (1999), *Providence and the Problem of Evil*, Oxford: Oxford University Press.

Taylor, S. E., and Thompson, S. C., (1982), "Stalking the Elusive Vividness Effect, " *Psychological Review* 89 (2), pp. 122-181.

Taylor, S., and Brown, J., (1988), "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103 (2), pp. 193–210.

——— (1994), "Positive Illusions and Well-Being Revisited: Separating Fact from Fiction," *Psychological Bulletin* 116 (1), pp. 21–27.

Thomasson, A. L., (2003), "Realism and Human Kinds," *Philosophy and Phenomenological Research* LXVII (3), pp. 580-609.

Thomson, J., (1971), "A Defense of Abortion," *Philosophy and Public Affairs* 1, pp. 47-66.

Tversky, A., and Kahneman, D., (1974), "Judgment under Uncertainty: Heuristics and Biases," *Science* 185, pp. 1124-1131.

van Fraassen, B. C., (1980), *The Scientific Image*, Oxford: Clarendon Press.

Wason, P., (1960), "On the Failure to Eliminate Hypotheses in a Conceptual Task," *Quarterly Journal of Experimental Psychology* 12, pp. 129-140.

Weatherson, B., (2003), "What Good are Counterexamples?" *Philosophical Studies* 115, pp. 1-31.

Weinberg, J. M., (2006), "What is Epistemology For? The Case for Neopragmatism in Normative Metaepistemology," in S. Hetherington, ed., *Epistemology Futures*, Oxford: Oxford University Press, 2006, pp. 26-47.

Weinberg, J. M., Nichols, S., and Stich, S., (2001), "Normativity and Epistemic Intuitions," *Philosophical Topics* 29, pp. 429-461.

Whyte, L. L., (1978), *The Unconscious Before Freud*, New York, NY: St. Martin's.

Williamson, T., (forthcoming), *The Philosophy of Philosophy*, Blackwell Publishing.

Wilson, T., (2002), *Strangers To Ourselves: Discovering the Adaptive Unconscious*, Cambridge, MA: Harvard University Press.

Wittgenstein, L., (1953), *Philosophical Investigations*, (trans. by G. E. M. Anscombe), Oxford: Basil Blackwell.

Zagzebski, L., (2004), "Epistemic Value Monism," in J. Greco, ed., *Ernest Sosa and His Critics*, Malden, MA: Blackwell Publishing, pp. 190-198.

ACTA PHILOSOPHICA GOTHOBURGENSIA
ISSN 0283-2380

Editors: Helge Malmgren, Christian Munthe, Ingmar Persson and Dag Westerståhl

Published by the Department of Philosophy of the University of Göteborg

Subscription to the series and orders for single volumes should be addressed to:
ACTA UNIVERSITATIS GOTHOBURGENSIS
Box 222, SE-405 30 Göteborg, Sweden

VOLUMES PUBLISHED

1. MATS FURBERG, THOMAS WETTERSTRÖM and CLAES ÅBERG (editors): Logic and Abstraction. Essays dedicated to Per Lindström on his fiftieth birthday. 1986. 347 pp.

2. STAFFAN CARLSHAMRE: Language and Time. An Attempt to Arrest the Thought of Jacques Derrida. 1986. 253 pp.

3. CLAES ÅBERG (editor): Cum Grano Salis. Essays dedicated to Dick A. R. Haglund. 1989. 263 pp.

4. ANDERS TOLLAND: Epistemological Relativism and Relativistic Epistemology. Richard Rorty and the possibility of a Philosophical Theory of Knowledge. 1991. 156 pp.

5. CLAES STRANNEGÅRD: Arithmetical realizations of modal formulas. 1997. 100 pp.

6. BENGT BRÜLDE: The Human Good. 1998. 490 pp.

7. EVA MARK: Självbilder och jagkonstitution. 1998. 236 pp.

8. MAY THORSETH: Legitimate and Illegitimate Paternalism in Polyethnic Conflicts. 1999. 214 pp.

9. CHRISTIAN MUNTHE: Pure Selection. The Ethics of Preimplantation Genetic Diagnosis and Choosing Children without Abortion. 1999. 310 pp.

10. JOHAN MÅRTENSSON: Subjunctive Conditionals and Time. A Defense of a Weak Classical Approach. 1999. 212 pp.

11. CLAUDIO M. TAMBURRINI: The 'Hand of God'? Essays in the Philosophy of Sports. 2000. 167 pp.