

Acceptance, Belief and Cognition

A. Max Jarvie

Department of Philosophy

McGill University, Montreal

December 2004

**A thesis submitted to McGill University
in partial fulfilment of the requirements of
the degree of Doctor of Philosophy**

© Max Jarvie 2004



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-494-12863-1

Our file Notre référence

ISBN: 0-494-12863-1

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

ABSTRACT

This is a study of a problem in the logic of belief revision. On the assumption of a number of fairly traditional views concerning the relationship between mind and world, the mechanics of perception, and the nature of belief, an argument is made to the effect that revision of extant beliefs is impossible even in the light of new perceptual experience. The argument turns on the ability of a cognitive system to recognize conflict among its thoughts and perceptions. A number of models of the mechanics of perceptual interpretation are explored, all of which are revealed to share a susceptibility to the problem as posed. Certain objections are taken up, the responses to which modify the scope of the original argument; although the problem may yet be said to arise in a number of crucial contexts where its presence is undesirable, some situations are found in which the problem can be dissolved. The problem is then reexamined in light of the epistemological position called *fallibilism*, with an eye to demonstrating that it arises notwithstanding the highly cautious perspective embodied in that position. A solution to the problem is then offered in the form of a family of model cognitive systems with certain properties. Because the problem is a feature of belief-based cognitive systems, the family of systems offered in arguing for a resolution of the problem is constructed on the notion that cognition, construed as information processing, normally proceeds without any epistemic evaluations being attached either to perceptions in particular or thoughts in general. The non-evaluative propositional attitude employed in normal cognition should, I argue, be what I call *acceptance*. The propositional attitude of belief, traditionally conceived of as occupying the role now given to acceptance, is accorded an extremely limited scope of application. Epistemic evaluation in general is itself restricted to contexts of decision only, its application arising only after conflicts or inconsistencies in the corpus of ideas are recognized.

RÉSUMÉ

Ce travail se penche sur un problème particulier à la logique de la modification des croyances. En base à un certain nombre de perspectives plutôt traditionnelles portant sur la relation entre l'entendement et le monde, sur le mécanisme de la perception et sur la nature de la croyance, nous postulons que la modification de croyances existantes est impossible, même à la lumière de nouvelles expériences perceptuelles. Notre argument se situe au niveau de l'aptitude d'un système cognitif à reconnaître les conflits parmi ses idées et ses perceptions. Nous analysons plusieurs modèles de mécanismes d'interprétation perceptuelle, lesquels partagent tous une susceptibilité au problème tel que formulé. Nous émettons certaines objections qui viennent modifier la portée de l'argument d'origine; bien que le problème survienne dans un certain nombre de contextes-clés où sa présence est indésirable, on trouve certaines situations où il peut être dissous. Nous réexaminons ensuite le problème depuis la position épistémologique de la faillibilisme, dans l'optique de démontrer qu'il survient malgré la perspective d'extrême prudence qui caractérise cette position. Nous proposons ensuite comme solution au problème une famille de modèles de systèmes cognitifs possédant certaines propriétés. Comme le problème est une caractéristique des systèmes cognitifs basés sur la croyance, la famille de systèmes que nous proposons en guise de solution découle de la notion voulant que la cognition perçue comme traitement de l'information se déroule normalement sans que l'on n'attribue d'évaluations épistémiques aux perceptions en particulier, ou aux idées en général. Nous postulons que l'acceptation est en réalité l'attitude propositionnelle non-évaluative qui survient dans une cognition normale. À l'attitude propositionnelle de la croyance, perçue traditionnellement dans le rôle actuellement attribué à l'acceptation, nous n'accordons qu'une portée d'application très limitée. L'évaluation épistémique en général est elle-même restreinte à des contextes de décision, son application survenant seulement après que l'on reconnaisse des conflits ou des inconséquences au sein du corpus d'idées.

TABLE OF CONTENTS

1. Introduction	1
2. Belief	13
3. Core Argument	42
4. Cognitive Systems: Ideals and Limitations	68
A. Compartmentalizing Cognition I	69
B. Compartmentalizing Cognition II	73
C. Cherniak's Long term Storage <i>qua</i> Filter	78
D. Cherniak's Short Term Storage <i>qua</i> Filter	80
E. Cherniak-Style Cognition & Conflict	85
5. Fallibilism	94
A. Analysis I: Strict Fallibilism	95
B. The Strict Fallibilist Escapes	98
C. Analysis II: Common Fallibilism	100
D. The Role of Belief in Common Fallibilism	102
E. Problems with Common Fallibilism	106
6. Evolution, Epistemology & Cognition	135
Division I: Evolution	135
A. Darwin, Teleology & Explanation	136
B. Persistence	143
Division II: Cognition	148
A. Evolution & Cognition	148
B. Evaluation & Cognition	153
C. Partial Systematization, Conflict, and the Will to Systematicity	165
D. Acceptance, Interpretation, Belief	176
7. Conclusion	191
Bibliography	205

Acknowledgements

Lest I commit any sin of omission in thanking those individuals who helped, either directly or indirectly, with the production of this work, I wish to thank first each of those persons whom I may have accidentally failed to mention in what follows. I give thanks to my mother and father, May and Ian Jarvie, for their support of my materially unproductive philosophical meanderings; to Valerie Guillet, for her excellent French; to McGill University, for its provision of material support and intellectual resources; to the Department of Philosophy of McGill University, for its astonishing tolerance of my extended dilettanteism; to the staff of that department, especially Paul Pietroski, Phil Buckley, Mario Bunge, Alison Laywine, Storrs McCall, David Norton, Sahotra Sarkar, and Sarah Stroud, for their instruction and advice; to my fellow students, both current and former, including Gerry Beaulieu, René Jagnow, Mohamed Mehdi, Dario Perinetti, and Brian van den Broek for fruitful intellectual engagement and conversation; to my supervisors, Jim McGilvray and David Davies, for their guidance, for their sincere support of the project despite their skepticism concerning some of its claims, and for their extremely generous commentary and criticism on the many drafts of this work; and to Nancy Zeligman, to whom I owe more than I can ever express.

Quot homines tot sententiae; suo quoque mos.

[So many men, so many opinions; his own a law to each.]

Terence, *Phormio*

Introduction

This essay explores a problem in the logic – and perhaps the phenomenology – of belief acquisition and revision. It does not seem that the particular problem raised here is very well explored in the literature; in fact, it appears not to have been raised at all, in the form I have chosen to raise it. Naturally, this raises the strong possibility that the problem is a pseudo-problem, discovered, so to speak, only by overlooking some simple and obvious fact or argument which would dispose of it at the outset. I have found no such philosophical oil of vitriol, and, moreover, cannot say whether the solution to the problem I offer in conclusion is itself genuinely sufficient to solve it.

The problem is relatively easy to state. Given some initial, fairly unremarkable assumptions concerning the relationship between mind and world, the mechanics of perception and cognition, sensitivity to skeptical arguments, and a fairly anodyne conception of the nature of belief, one may derive the following conclusion: that rational revision of existing beliefs in the face of new evidence is impossible. Clearly, we revise opinions that we hold on various matters which relate to and are responsive to evidence on a regular basis; that is to say, we rationally revise propositions we hold and act upon *all the time*. Hence the problem. Either there is something wrong with the collection of assumptions, or the logic of the argument leading to the claim of unrevisability is faulty.

In what follows, I offer as a solution to the problem the idea that human cognition is not grounded in beliefs as traditionally construed. If this solution is correct, it indicates that the term ‘belief’ as generally used both in philosophical literature and in everyday discourse actually conflates (at least) two quite crucially different concepts: the traditional understanding of belief as a truth-ascribing propositional attitude and another concept I have chosen to name *acceptance*, a propositional attitude which ascribes no epistemic evaluative attributes to its objects, for example, the attributes ‘held true’ or ‘held plausible’, at all.

Some comments should be made regarding the overall perspective from which the paper may be said to operate, as well as the presuppositions that guided the generation of the problem and which also furnished constraints for the suggested solution.

The course of investigation undertaken in this project is substantially guided by concerns which may be described as Humean, if not in scope or substance, then at least in spirit. Hume was no champion of radical skepticism; he is best understood as a post-skeptical philosopher. In the words of David Fate Norton: “Hume supposed (a) that the Cartesians (especially Malebranche) and Locke and Berkeley had...taken traditional metaphysics and epistemology to its skeptical conclusions; (b) that these skeptical conclusions had been soundly and validly established; and (c) that the most important remaining task of philosophy, given these well-established and obvious conclusions, was to show how we are to get on with our lives, particularly our intellectual lives.”¹ Like Hume, I judge the skeptical arguments to be found in the philosophical literature to be entirely valid and sound, and have found no plausible argument against them. In the face

¹ Norton, David F., *An Introduction to Hume's Thought*, in the Cambridge Companion to Hume, Cambridge University Press, © 1993, p. 5.

of the apparently insurmountable wall of doubt that lies – as such arguments conclude – between ourselves and knowledge of an external world, it is worth asking how we manage to operate in such conditions and where our methods will take us. We know that Hume's answer to this question was to build a 'new science of the mind' to describe our cognitive powers and that his deliberations led him to conclude that our acquisition of those beliefs which go beyond observational and mathematical (and logical) facts arises from the arational workings of human psychology. His conclusions aside, it is of interest that his project, this 'new science', took as its fundamental elements the immediate objects of thought; the Humean perspective is radically subjectivist.² Despite having read and thought about some of the great quantity of philosophical literature produced in the twentieth century which sought to push mentation and meaning outward from the individual cognitive subject onto the world and the linguistic society of humankind, I think Hume's starting point the most intuitively satisfying; it proposes that we seek a livable skepticism.³

This feature of the Humean project is best understood, I think, by reflecting on Hume's observations regarding his variable attitude towards philosophical (in particular, skeptical) reasoning when placed in different circumstances: on the one hand, within the confines of his study, and on the other, while having dinner or engaging in a game of backgammon with his friends.⁴ In the passage from the *Treatise* in which he remarks

² By which it should be understood that Hume's 'new science' took as its starting point the perspective of a thinking human subject, not David Hume in particular.

³ And may well go beyond this. Hume's suggested description of our cognitive powers and mechanics of mind may conceivably be applied to individual cognitive agents devoid of skeptical worries of any sort. Being no expert on the writings of Hume, I would not want to impute this particular thought to him; however, this suggested application is crucial for any description purporting to accurately map the mental life of all persons – even those of a positively antiskeptical bent.

⁴ Hume, David, *A Treatise of Human Nature*, P. H. Nidditch (ed.), Oxford University Press, 1978, pp. 263-274, especially pp. 268-270.

upon this contrast, Hume notes that his engagement in ordinary activities “obliterates” the skeptical gloom brought on by his philosophical reflections, and that upon returning to his study his earlier skeptical concerns seem “cold” and “strain’d”.⁵ Yet it seems implausible that Hume would have thought this contrast worth remarking upon if this shift in his attitude had happened only once. What makes it of philosophical interest, in a way that resonates with our own experience of encounters with skeptical thought, is that our attitudes may vacillate between the two extremes depending upon whether we find ourselves in the study or the street; moreover, the entertainment and even the acceptance of skeptical arguments does not seem to substantially affect us when we sit at the table to eat, or roll the dice for our next move in the game. There is a philosophical puzzle here; and we may understand Hume’s project in part as an attempt to explain - or somehow make sense of - how such vacillations are possible, and how it is that taking skepticism seriously does not necessarily lead to cognitive paralysis. The initial articulation of the problem which is the primary concern of this work does not itself require that we pay any attention to this particular Humean concern; however, the exploration of the problem and the solution offered in the latter stages of the investigation is sensitive to the seriousness of the puzzle just described.

Hence this paper concerns itself exclusively with the relationships which may be found among those immediate objects of thought: perceptions, ideas, opinions, beliefs. The efforts in the contemporary period to frame meaning and mind in more objective terms seem to spring from (perhaps among other things) anxieties about the possibility of real communication among thinking individuals through language, which anxieties I do not share. With respect to this issue (and *pace*, for the moment, skepticism about the

⁵ Ibid., p. 269.

external world and other minds), I think that Chomsky may be right: there may well be certain innate structures that human cognitive systems share, and these structures may explain why our attempts to communicate with each other, by and large, meet with apparent success (at the very least, practical success). However, real communication, the shared apprehension of some objective meaning to which our words are supposed to attach, may be no more than an expeditious fiction. Moreover, again as Chomsky has noted, we may individually be possessed of some innate concepts, and many if not all of our complex concepts may arise through the composition of these simple concepts into complexes. I reserve judgment on the question of whether any of this is actually correct, but it seems a much more plausible starting point for untangling the issues concerning the operation of human cognition than either objectivist or intersubjective accounts of the formation of concepts and the origin of meaning, insofar as such a nativist approach may be blended painlessly with skepticism concerning the external world; most importantly, it does not seem that all the avenues of solution which begin with these skeptical intuitions have been exhausted. That offered in this paper is one such unexplored avenue.

The central problem of the paper may not require such a skeptical perspective for its generation; but my adoption of this viewpoint may go some way towards explaining why I have chosen to follow the path taken here in exploring and resolving it. For example, it may help to explain why there is no discussion of externalist theories of justification, which may after all provide a way to resolve the issues explored here. It may also help to explain the rather introspective nature of the explorations themselves.

Apart from the presuppositions mentioned above, this paper is also guided by the notion that the act of perception is an interpretative process that conducts its operations

upon the raw data delivered to the mind by the senses. I subscribe to some form of what McDowell calls the 'highest common factor' theory of perception to which skeptics are particularly prone. The problem which this paper concerns itself with is without question dependent upon such a view of perception; if a theory of direct perception turns out to be the correct way of understanding perception, then the worries of this essay might easily dissolve. However, the problem which this essay explores is not intended as preparation towards an argument for any theory of direct perception. The solution offered, in the latter stages of the discussion, adheres to the presupposition that some sort of representational theory is correct.

It is not intended, however, that the articulation of the problem depend upon a representational theory of perception which makes use of any particular thesis regarding the ontology of sense-data. The non-propositional contents which the 'raw data', referred to in what follows, consist in *may* be present to the consciousness of a cognitive system; but they are not understood here as anything more than undifferentiated, unconceptualized sensations until they enter the interpretative process. The process, or at least some parts thereof, of the formation of a perception which the processing of such raw data yields may itself be open to inspection by consciousness, although in most cases it likely moves too quickly to be properly apprehended. The result of such processing is a fully-formed perception, which may well mean, in any particular case, perception as of an object external to ourselves.

Perhaps most important of all the presuppositions of this paper is the low status afforded to any part of the justificationist framework which has guided most of the discussion in epistemology and scientific method in the modern period. As with the

presupposition concerning the mechanics of human perception noted above, this presupposition seems to fall naturally from the Humean outlook adopted overall. If certain knowledge concerning the external world is not available to us, the project of looking for *justifications* for our beliefs seems a misguided one. What is wanted is an explanation of, on the one hand, how we come to hold what we do hold, and on the other, what makes our holding rational; but this problem is a very general one, to which the construction of a logic of justification framed in terms of *epistemic warrant* is merely one response. An occasional mention in what follows of the *license* or *warrant* for what we accept, afforded us by the procedures we follow, should not be interpreted as an acquiescence in the justificationist project; all that is intended is a tip of the hat, an acknowledgment of the concerns of those who are so guided. That being said, we can allow that there may be a place in the description of human cognition for justification, just as there may also be a place for probabilistic reasoning; but their positions must be ancillary, not central, as so much of the literature has supposed. The exploration to follow may hint, in different places, at the sort of position I have in mind; a brief discussion of this point will be reserved for the conclusion.

Last on the menu of presuppositions is the utility of the concept of *belief* itself. The generation of the problem requires a notion of belief understood as a truth-taking attitude. It is of interest that some languages possess no analogue of the word ‘belief’: Hebrew, both ancient and modern, for example, does not capture this concept within its linguistic resources. In those places where we might expect to find the word ‘belief’ or ‘believes’, Hebrew equivalents of the words ‘thought’, ‘think’, and ‘faith’ stand in their stead. Such examples alone suggest that the concept of belief as it is used in romance

languages and in English may be a conflation of a number of different, albeit related, concepts. Because of this uncertain status that the word enjoys as a possible topic of philosophical deliberation, and because of the central importance it has in the essay as a whole, the merit of this particular presupposition is addressed at length in the first chapter. The other presuppositions, as will become clear, are left purely as such and no attempt is made to defend them.

The essay is written in the spirit of a piece of investigative journalism, a perhaps unorthodox approach to the composition of a philosophical monograph. Given the nature of the problem and the paucity of material available in the literature that appears to address it, this seemed, on balance, the most reasonable approach. The structure of the work may be broken down as follows.

As the paper is focused on the nature of belief, we begin with a territorial survey of considered opinion on the subject of the term. The goal in this section is not to exhaust all extant opinion on the meaning of the word; we seek only to discover whether there is anything approaching a settlement on these matters. As it turns out, unsurprisingly, there is none; what may be drawn, however, from the consideration of the various points discussed is that there is some consensus to the effect that discussion of the concept of belief is not at present an illegitimate or wasted exercise. Despite there being no agreement on the underlying meaning of the term, it seems largely agreed that exploration of the various meanings both connotative and denotative which may be attached to *belief*, as well as the consequences of those meanings, may serve at minimum to specify how we may come to eliminate it or determine its role in cognition. Exploring

its possible meanings and the consequences of those meanings, then, is a reasonable endeavour.

The discussion then proceeds to a description of the problem itself and some of the assumptions which appear to motivate it. The problem focuses on the treatment of belief as a truth-regarding attitude in conjunction with a view in which perception is regarded as the interpretation of largely unformed data. It is noted that the possibility of rationally revising one's beliefs requires a particular motivating factor: the presence of a conflicting belief within the cognitive system. Although the persistence of such a situation may be intolerable, such a condition must arise, if only briefly, in order for a conflict to be recognized that would prompt reconsideration of an extant belief or collection thereof. In many cases, the conflicting belief that would motivate such a reconsideration will be a perceptual belief. The problem arises because, given the influence that extant beliefs may have over the interpretation of raw information into fully-formed perceptual episodes, it is not immediately obvious that any potentially conflicting interpretation may come to be believed; the extant beliefs of the cognitive system have a clear potential to interfere with the very formulation of an interpretation of experience which would yield a new belief that conflicted with one already present in the system. This influence of extant beliefs in the act of perceiving is explored and a variety of forms that that influence may take are analyzed for their consequences. In addition to exploring the consequences of interpretation in perception by the employment of a conception of belief as an *absolute* truth-regarding attitude, a brief examination is made of how the story might play out if a probabilistic understanding of belief is employed. The tentative conclusion reached is that the problem emerges no matter which form of

influence is chosen. The notion of *acceptance* is mentioned briefly, but is neither fleshed out nor dwelt upon here. Instead, the discussion turns to refinement of the problem through the examination of a few serious objections concerning its formulation.

The first objection takes to task the clearly Quinean viewpoint upon which the simple form of the problem depends. If Quinean mental holism accurately portrays the workings of the human cognitive system, the problem manifests itself very quickly. The adoption of a cognitive model which allowed for the sharply delineated contextualization of thought suggests a way of solving the problem, so the objection takes the form of rejecting the holistic model in favour of such a contextualized model. The very idea of such a model is briefly explored here, and some of its possibly undesirable consequences noted; discussion then turns to a particularly promising and well-described theory of context-bound cognition. The minimal rationality model of Christopher Cherniak, with its attendant compartmentalization of cognition, provides for the possibility of achieving the crucial cognitive condition in which two conflicting beliefs may simultaneously come to the attention of the conscious mind. Cherniak's views, and the consequences thereof, are examined with this objection in mind. The exploration of this material yields some surprising results. In some cases, a Cherniak-type cognitive model does allow for the possibility of two conflicting beliefs to be present to the conscious mind; this allows us to admit the possibility that a certain type of cognitive system grounded in belief may sometimes achieve the cognitive condition necessary for rational reconsideration (and change) of opinion. However, in a great many cases in which we would want some such rational reconsideration to be possible, Cherniak's model is shown to suffer from the same limitations as holistic models with respect to the problem. We conclude that while

adoption of a Cherniak-type model of cognition may represent a significant advance on Quinean holism, both with respect to the dissolution of the central problem and for other reasons, it fails as a global solution to the problem. Adoption of Cherniak's model as a partial solution to the problem also introduces a surprising consequence insofar as it requires us to reject Quinean holism not only as a genuine description of human cognition, but also as its ideal form.

Discussion then turns to the matter of the epistemological position of fallibilism, with an eye to determining what impact the problem has on a cognitive system which may be described as conscious of its own fallibility. We begin by attempting to clarify the meaning which is attached to the term, both by exploring its historical origins in the literature and some contemporary definitions. In the course of this exploration, a curious division emerges between, on the one hand, the meanings originally intended by C.S. Peirce and Karl Popper, and on the other, that which appears to be the word's current meaning as an epistemological term of art. The current meaning, insofar as it can be extracted from some contemporary definitions and discussion, is then further examined for its stability; under scrutiny, it appears quite prone to collapse into a form more closely aligned with the meaning intended by its progenitors. The significance of the disagreement lies in the fact that the fallibilisms of Peirce and Popper do not appear to be susceptible to the problem as stated at all; the seriousness is compounded both by the fact that Popper's overall outlook is anathema to most contemporary philosophers, and that the position of Peirce is in crucial ways all too similar to Popper's on the matter concerned. Because of this, realignment of the contemporary definition of fallibilism with the earlier intended sense of the word is no easy option for contemporary epistemologists

to embrace. In deference to the possibility that the contemporary sense of the term ‘fallibilism’ is in fact more stable than the pokings and proddings of our discussion have shown, we proceed on the assumption of its stability and explore its susceptibility to the central problem. The conclusion reached, through rehearsal of some of the argument laid out in the chapter concerned with detailing the structure of the central problem, is that this modern understanding of fallibilism is indeed vulnerable to the problem and that no solution appears to be available.

Our discussion of fallibilism concludes our excursion into major objections which may be raised against the cogency of the central problem. The final discussion turns instead to the construction of a positive account which may solve the problem; this requires the specification of a cognitive system which is capable of cognition of the sort normally associated with human minds, yet which is not grounded in belief. We begin by exploring a topic somewhat removed from the issues at hand: the Darwinian theory of evolution, and its possible analogical extension into the domain of epistemology. This is old territory, to be sure; but some attempt is made to draw the focus of the discussion to a somewhat neglected consequence of the theory of evolution concerning the *persistence* rather than the *change* of genetic traits over a span of generations. This insight is then applied to the construction of a model for a cognitive system which respects the constraints of an evolutionary epistemological outlook, and the consequences explored. Whether or not this model succeeds in overcoming the central problem, I think it is an illuminating alternative worth consideration.

Belief

When we first read Descartes' *Meditations*, we respond to it in different ways. With Descartes leading us on by the light of a melting candle, we methodically descend a staircase of doubt into the cellar of the house of knowledge. As we leave behind the comfort of the upper rooms, our surroundings become increasingly indistinct; darkness closes in around us until we cannot make out the forms of our own bodies, not even our own two hands. Having reached bottom, the candle guttering, Descartes strikes a match, finds a light switch, and floods the immediate area with light; he then shows us how to build a ladder to take ourselves back out.

Most people find the latter part of the *Meditations* largely unconvincing: the method for building the ladder is not as sound as it first appears. Our criticism of the tools and materials provided for us is sharpened by our appreciation of the stepwise process which brought us to this low place. Despite Descartes' explicit emphasis on the *methodological* nature of his arguments, those arguments resist resolution with almost punitive force. Although we may agree with Descartes that we ultimately accept the claims that we do with reason, the difficulty of formulating a solution compatible with his skeptical arguments can provoke serious philosophical anxiety. As Hookway says in his digest on Quine:

...we feel the force of the possibility that all might be a dream; and students in their first weeks of philosophy are readily convinced of the need to reply to Descartes' challenge.¹

¹ Hookway, Christopher, *Quine*, p. 198.

Most reply to the challenge by recognizing the dire consequences of the Cartesian method: once (almost) everything in the edifice of knowledge has been opened to doubt, nothing can be relied upon; the possibility of epistemic paralysis looms large. Recoiling from this, most protophilosophers set about the task of finding ways to escape the tight confines of the skeptical well, Descartes' own arguments having been long discredited by various eminent persons. We need not recount the sad litany of frustrated attempts here.

Others react by accepting the consequences of the skeptical arguments, either by embracing those consequences or simply resigning themselves to them. Hume's acquiescence in arational psychology as the foundation for our attempts to acquire knowledge could be looked upon as an example of resignation; Karl Popper's propounding a doctrine of falsificationism, as an example of wholehearted partnership with the skeptic.

The anatomy of our reactions to the methodical skepticism so beautifully described in the *Meditations* finds its focus in the notion of belief. The force of the argument is felt as a challenge not only to knowledge and the possibility of knowledge, but also to our sense of entitlement with respect to what we believe. For, while we may aspire to knowledge (and may even possess some), where knowledge is construed in the traditional mode of *justified true belief*, what we usually (or more plausibly) take ourselves to have is no more than reasonable belief. It is commonly understood that on the basis of such beliefs we make our decisions, construct predictions, and assert our opinions. Descartes' skeptical arguments might not be felt so strongly if they merely indicated that our methods of inquiry, and ratiocination generally, were not *completely*

reliable. If they suggested only that we have a tendency to go wide of the mark, or employ misleading taxonomies of objects, processes, and forces, we might take those arguments as merely a call to vigilance and the close inspection of our claims. However, in the eyes of most readers, the Cartesian arguments go much farther than this. They suggest, in the strongest possible terms, that we could be utterly and completely wrong about virtually everything we claim to have reason to believe, and that we inhabit a sensory world of such poverty that we may never be able to realize the extent of our errors. Our first reflections upon these arguments can lead to consideration of the possibility that we may have no reasonable beliefs at all: that what we think of as justifications for those beliefs are wholly without merit, being founded in the unreliable senses or the error-prone and largely irrational workings of the mind.

Before examining the possibility of extricating ourselves from these worries, it is worth turning first to the notion of belief. What is a belief? A legitimate entity, or an empty shell? If there are “no such things” as beliefs (or other mental entities), or if their ontological status cannot be decided, is it legitimate for us to discuss them? These and related questions have received many answers, and mapping the terrain in full would be a daunting task. Even restricting ourselves to the modern period, we will have time only for a perfunctory tour; nonetheless, we may find sufficient answers with which to equip ourselves for assaulting the main problem of the essay.

As in so many other areas of philosophical inquiry in the analytic tradition, Quine can serve us here as useful sounding-board for determining what, if anything, a belief *is*. In the opening to the essay *The Web of Belief*, Quine and his coauthor Ullian propose that

...[m]ost of us believe that Hannibal crossed the Alps, that Neptune is a planet, that frozen foods thaw when left at room temperature overnight. We also share beliefs of a higher order – beliefs about our beliefs...we agree that what we think we see is, much more often than not, genuinely there. Seeing is not quite believing, but it, together with the continual “testimony” of our other senses, fairly bombards us with new material that requires assimilation in our body of belief. So it is that each of us is continually adopting new beliefs, rejecting old ones, and questioning still others. One’s repertoire of beliefs changes at least slightly nearly every waking moment, since the merest chirp of a bird or chug of a passing motor, when recognized as such, adds a belief – however trivial and temporary – to our fluctuating store.²

The conclusions that can be derived from this initial paragraph are (as intended, one can be sure) largely indisputable. One can, and some do, take issue with the notion that ‘seeing is not quite believing’, but by and large the notion that our senses provide us with an unmediated, accurate apprehension of reality external to ourselves is not a popular one. Surely it is not a notion that Quine and Ullian favour, as they state that the effects of external events upon our senses only become beliefs when they have been *recognized* as being indicative of a particular kind of event – the chirps or chugs of birds or motors, to use their own example. We all have a repertoire of beliefs; in *The Web of Belief*, it eventually transpires that this repertoire actually constitutes a self-supporting web of coherent beliefs, which is both continually altered by new input and itself acts as a mechanism for interpreting such input.

What of the individual constituents of this web? The passage continues:

And yet, for all the liveliness of fluctuation of beliefs, believing is not itself an activity. It is not like scansion or long division. Nor is it a fit or mood, like joy or grief or astonishment. It is not something that we feel while it lasts. Rather, believing is a disposition that can linger latent and unobserved. It is a disposition to respond in certain ways when the appropriate issue arises. To believe that Hannibal crossed the Alps is to be disposed, among other things, to say “Yes” when asked...Some beliefs, like the one about Hannibal, we shall probably retain while we live. Some, like our belief in the

² Quine, Ullian, *The Web of Belief*, 1st edition, p.3.

dependability of the neighborhood cobbler, we may abandon tomorrow in the face of adverse evidence. And some, like the belief that a bird chirped within earshot, will simply die of unimportance forthwith.³

Despite the characteristically sparkling prose, the characterization of belief here offered is not quite as clear as one might hope it would be. Quine and Ullian, like Ryle before them (although for different reasons), hope to restrict beliefs to the domain of dispositions to behave, which is perhaps too radical. It is not to be disputed that many beliefs do fall into this category; much of what we believe is not held constantly before the conscious mind, or at least, not held within the frame of immediate attention that our minds are capable of, but rather sinks into a background where their presence may influence our perception and other forms of judgement. In this way, most beliefs may be like Husserl's phenomena of consciousness, inhabiting the regions between the focus of attention and the distant mental horizon, or possibly like Christopher Cherniak's items of information stored in long-term memory compartments. However, when we are prompted by some appropriate stimulus, we call individual beliefs into our frame of attention (or our so-called 'short term memory'), as we might when someone asks us whether we believe that Hannibal crossed the Alps or if we heard a car backfire just now. When we do so, and are disposed to act on the basis of such beliefs (as we might by responding "Yes" or "No" to an inquiry, as appropriate), it no longer seems correct to refer to the belief as merely a disposition; it has become, in some way, activated or occurrent. In such circumstances it would be fair to characterize an individual as, for example, *thinking that* or *taking it to be true that* Hannibal crossed the Alps. To be fair to Quine and Ullian, this intuitive division of beliefs into dispositional (stored, latent, unobserved by the

³ Ibid., p.3-4.

individual's conscious mind) and occurrent (actively thinking that or believing that Hannibal crossed the Alps), is one that they do not share. The web of belief is a continuous whole, and is exposed *as a whole* to the impact of experience upon it. There is no separate arena to be characterized as a place in which a particular belief can be brought before a focus of attention in which that belief can be considered, its consequences and relationships with other beliefs or experience lately acquired to be there examined and adjusted. Despite this, Quine and Ullian do appear to pay token attention to the possibility of this intuition, when they raise the following point:

We often talk as if believing were something that a man does *to* something: to some intangible thing which is *what* he believes. To name this thing we enlist a sentence as subordinate clause. For example, we speak of the belief *that* Hannibal crossed that Alps and *that* Neptune is a planet. We use a sentence, with "that" prefixed, as a name of the "thing" believed. Now what manner of thing is this believed thing – *that* Hannibal crossed the Alps? To say that it is just the sentence itself seems mistaken... What then are they?

This, like various other philosophical questions, is better deflected than met head on. Instead of worrying about the simple verb "believes" as relating men to some manner of believed things, we can retreat to the word-pair "believes true" as relating men directly to sentences. We can retreat to this without claiming that believed things are sentences; we can simply waive that claim, and the philosophical question behind it. After all, our factual interest in what some speaker of English believes is fully satisfied by finding out what sentences he believes to be true.⁴

Although this elaboration is intended to render the notion of belief more pellucid, in the absence of certain knowledge about Quine's outlook as expressed in other essays, it has rather the opposite effect. Quine's perspective on propositional attitudes is coloured by his verificationist tendencies. In examining the intuition that believing is an act ("We often talk as if believing were something that a man does *to* something"), which characterization he and Ullian have earlier denied ("believing is not itself an activity"),

⁴ Ibid., p. 4-5.

the passage quoted directly above proposes to deflect or deny that there is anything at issue here, by heading off any intuitions one might have about what the objects of such supposed acts of belief actually are. We are to be satisfied with the answer to the question of what a speaker believes, which can be answered by determining what sentences a speaker believes true. If we assume, in the process of querying an individual, that we can verify assent, then we have done all that we can do. Because of Quine's allegiance to his thesis of the indeterminacy of radical translation, there is for him no fact of the matter about what the contents of a speaker's sentences are, that is, what these sentences may mean; consequently there also cannot be any fact of the matter as to what the contents of the speaker's beliefs may be, on the basis of which the speaker assents to or denies our queries. All that can be verified is assent, and so there ought to be no philosophical worries about what the 'thing' is to which the speaker is assenting; it is only our usual, uncritical usage that suggests otherwise.⁵ It is worth bearing in mind this theoretical baggage when interpreting the above, as otherwise the analysis seems to leave belief as a kind of black box, the question of the ontology of belief being dismissed as not relevant to the goals of determining what a speaker believes. Answering the question of what a belief is, is as much part of Quine and Ullian's task here, as their factual interest in what some speaker of a language believes or does not believe. From their perspective, the answer given is perfectly adequate to address the ontological question: because there is no fact of the matter as to what a belief is about, we can completely analyze an individual's believing that Hannibal crossed the Alps as that individual's being disposed to respond in certain ways with respect to stimuli which somehow relate to the sentence

⁵ It is not possible, unfortunately, to verify whether Ullian is in complete agreement with Quine's views on this matter, as they are not made explicit here, and there is very little other work by Ullian which might be used to triangulate his own thoughts; so we will assume agreement here.

“Hannibal crossed the Alps”. One manifestation of this disposition is to utter “the sentence ‘Hannibal crossed the Alps’ is true” when appropriately stimulated.

There is, however, something disquieting about this analysis, in that it appears to rest on the idea of *disposition* as a primitive. Just as the disposition called *solubility* depends on the structure of a particular physical substance, one wonders what structure might underlie the disposition to believe. There is an independent debate about dispositions, which focuses on this very topic of their suitability for a grounding role in any area of discourse. We might have called upon such a debate here, as a motivation for seeking what might qualify (if anything) as providing a ground for dispositions. As it turns out, though, we have a motivation to do so which is internal to Quine’s (if, perhaps, not Ullian’s) outlook. Although the topic does not arise in the course of his discussion at the opening of *The Web of Belief*, it does arise in other areas of Quine’s work. Although Quine and Ullian choose to characterize belief as a kind of behavioural disposition, Quine elsewhere states that

Mental states and events do not reduce to behaviour, nor are they explained by behaviour. They are explained by neurology, when they are explained. But their behavioural adjuncts serve to specify them objectively. When we talk of mental states or events subject to behavioural criteria, we can rest assured that we are not just bandying words; there is a physical fact of the matter, a fact ultimately of elementary physical states.⁶

So it appears that Quine himself is not content to leave dispositions to behave in certain ways as some fundamental primitive in the explanation of mental states, such as belief. It is not clear whether we should draw any conclusions about his failure to even mention this point within the confines of the essay *The Web of Belief*, since he and Ullian

⁶ Quine, W.V.O., “*Facts of the Matter*”. *Essays on the Philosophy of WV Quine*, p. 167.

do not make a point of claiming that dispositions are primitives. The discussion in that paper may invite this interpretation but does not imply it. Perhaps it is a point upon which Quine and Ullian agreed to disagree, allowing their readers to fill the silence with their own conclusions.

In addition to having this internal motivation to go beyond behavioural dispositions in the analysis of the term belief, Quine also has the next step already laid out: behavioural dispositions must be cashed out in terms of the physical science of neurology. We are now in a position to see how Quine (if not Quine and Ullian) answers the question of what a belief is. Initially characterized as a behavioural disposition, a belief is ultimately to be explained as a physical state of an organism. Note that there is nothing in this outlook which precludes the possibility of multiple realizations for a particular belief; given two individuals who assent to sentences of exactly the same form, there will be a physical fact of the matter about each individual which explains the behavioural disposition we have identified each as possessing; while we have identified the same disposition in each individual, the physical facts which explain their dispositions may be different in each case. Moreover, the characterization in terms of behavioural disposition has little explanatory value; it is only a convenience which allows us, on a case-by-case basis, to specify objectively what a particular belief amounts to. The objective specification in terms of dispositions to behave can presumably then be resolved into specifications at the neurophysiological level, identifying the physical structures which ground those dispositions. This is clearly of a piece with Quine's general outlook, which finds no explanatory value in any form of discourse outside of the physical sciences.

Thus for Quine, the proposal that to believe is to be *disposed to behave* in certain ways when prompted has as little explanatory value as the proposal that to believe is to perform some act, that believing is something someone *does* to some *thing*, such as *regard* some proposition or sentence as true; however, the former proposal has the miniscule merit that it goes some way towards an objective specification of the term which might usefully point the way to identifying physical structures which explain the dispositions, whereas the latter cannot even hope to do that. This is the real reason that Quine would prefer the dispositional analysis over an analysis in terms of believing being an act, in particular a private or unobservable act of taking an attitude towards a proposition. He would prefer it, in other words, not because he wishes to protect the dispositional analysis from the possible introduction of non-dispositional notions like acts and attitudes, but because the dispositional analysis shows promise in helping us reach a real explanation of the life of the mind in terms of physical states and objects, and the other does not. However, generally speaking, neither proposed analysis has any real value at all.

Having used Quine and Ullian to shed some light on the terrain, we may be in a position to map a portion of it. To begin with, it should be mentioned that there are two issues at hand: an ontological question as to what beliefs are, and a semantic question as to what talk of beliefs (and other propositional attitudes) amounts to. Quine's answer to the ontological question essentially appears to be eliminative; it is not entirely clear how he answers the semantic question. There is an eliminative tendency here as well, as when he disparages the explanatory value of discourse that makes use of propositional attitude vocabulary; however, he does not directly address the question of eliminating such

discourse altogether. Perhaps here there is a little bit of Carnap's distinction between formal and material modes left in Quine's outlook, in particular Carnap's permissiveness with respect to linguistic forms, his *Principle of Tolerance*.⁷ Perhaps also, Quine may regard his analysis in the manner of Ryle: on the one hand, he is analyzing the usage of propositional attitude vocabulary, and on the other, helping to clear up confusion (especially ontological confusion) which our normal use of that vocabulary engenders.

In mapping the terrain, we can also find a place for Ryle. Ryle is notoriously associated with providing answers to the semantic question only, and as just mentioned, regarded his own task as helping to remove the mental confusion (in particular, category errors) which he saw as pervading propositional attitude discourse. Rather than eliminate mentalistic vocabulary, Ryle sought to rework our understanding of the terms of that discourse. How much this commits him to behaviourism, with which he is also notoriously associated, is unclear; as such, his ontological position on the mental is also unclear. Applying his notion of category mistakes to the mind, he argues that we are easily misled into thinking that the mind is a special kind of thing which occupies a place in the world, which manifests its presence through the behaviour of individuals. Instead, we are invited to treat the mind and everything associated with it – beliefs, desires, and so on – as consisting in or definable in terms of these manifest behaviours, or patterns of them. However, Ryle never gave an account of behaviour itself in fully non-mental, physicalistic terms; behaviour is always treated as fully intentional. Ryle's outlook is still a species of behaviourism, but it is not reductive or eliminativist in either the manner of Quine, or that of the behavioural psychologists Watson and Skinner. Hence there is no

⁷ "Let us be cautious in making assertions and critical in examining them, but let us be tolerant in permitting linguistic forms". Carnap, R., *Meaning and Necessity*, 1956, p. 221.

clear ontological position to be found in Ryle's work; as a consequence he is considered to have contributed only to the semantic side of the debate. The upshot of that contribution, for our purposes, is that he expresses no wish to abandon mentalistic discourse, and hopes only to clarify it.

Among contemporary philosophers, Daniel Dennett can also be mapped as contributing largely to the semantic side. Dennett regards his own views as a function of those of Quine and Ryle, and he manages to blend the two nicely. Dennett expresses an eliminativist strain where ontology is concerned; he says that he has tried to undermine two strands of realism which he has found philosophers to be susceptible to:

- (1) Realism about the entities purportedly described by our everyday mentalistic discourse – what I dubbed folk psychology – such as beliefs, desires, pains, the self;
- (2) realism about content itself – the idea that there have to be events or entities that *really* have intentionality (as opposed to the events and entities that only behave *as if* they had intentionality).⁸

Dennett, like Quine, is a scientistic physicalist; he regards living creatures as 'composed of thousands of deliciously complicated [biological] gadgets, yoked together opportunistically but elegantly into robust, self-protective systems'.⁹ Sentient creatures, in particular, can be regarded as amenable to a particular sort of analysis, which Dennett terms 'the intentional stance'. The stance is, as he puts it, a 'tactic'; it is an heuristic which consists of the presupposition that an entity is 'an approximation of the ideal of an optimally designed (i.e. rational) self-regarding agent'.¹⁰ This presupposition carries no ontological weight, and is introduced purely to provide leverage for generating

⁸ Dennett, Daniel C., "Self Portrait", in S. Guttenplan (ed.), *Blackwell Companion to the Philosophy of Mind*, pp. 239-240.

⁹ *Ibid.*, p. 239.

¹⁰ *Ibid.*

predictions of an entity's behaviour. Dennett contrasts his 'intentional stance' with the 'design stance' and the 'physical stance', and suggests that our motivation for adopting the 'highest' stance arises in cases in which the design stance and the physical stance both prove to be impractical, often (but not exclusively) for reasons of complexity.¹¹ There are no facts of the matter where beliefs and other propositional attitudes are concerned. Like Quine, Dennett finds facts only at the neurophysiological level, where one might speak of function and design, or the physical level, where one can speak of atomic and subatomic interactions.

Assuming that the above telegraphic restatement of Dennett's views is accurate, he can be regarded as more a product of Quine than of Ryle. The physicalism, the denial of any ontic status to beliefs and other propositional attitudes, and the allegiance to science all figure prominently in Dennett's outlook. However, he finds a clear purpose in propositional attitude talk which goes beyond that of Quine. Whereas Quine wants to reduce propositional attitudes to (purely physical) behavioural dispositions, and use those dispositions as objective specifications to be given over to neuroscience and linked with brain function, Dennett sees propositional attitude ascription as a practical foundation for certain kinds of prediction. As far as what that sort of talk amounts to, above and beyond this practical purpose, Dennett does fall in, ultimately, with Quine: insofar as propositional attitude talk depends on intentionality, it must be cashed out in the vocabularies of the design stance and the physical stance. Having taken out a loan of intelligence, rationality, and intention in the pursuit of predictive power, the debt is repaid ultimately through decomposition of an intelligent intentional system 'into hierarchically

¹¹ Dennett, Daniel C., "Intentional Systems", in John Haugeland (ed.), *Mind Design*, p. 225.

structured teams of ever more stupid intentional systems, ultimately discharging all intelligence-debts in a fabric of stupid mechanisms'.¹²

What Dennett draws from Ryle is his desire to untie knots of conceptual confusion. Our employment of propositional attitude talk invites such confusion, in Dennett's eyes, and encourages us to reify such purported entities as beliefs and desires. Propositional attitude talk has meaning, but does not refer to the entities it appears to refer to. Rather, the meaning of such discourse is tied to its practical purpose as a predictive tool, and thereby only indirectly to physical entities which the predictive tool may help us ultimately identify. Because those physical entities and their organization must be settled on a case-by-case basis, depending on whether we are investigating the structures of humans, cats, martians, or discrete state machines, one might think that from Dennett's perspective the ultimate semantic value of propositional attitude discourse could be viewed as cashing out into an increasingly enfeebled and uninteresting disjunction of physical attributes and relations. For Dennett, however, the intentional stance is a crude and approximating tool. The adoption of the intentional stance may *suggest* norms to which the physical makeup of an individual creature must be responsible; but the physical makeup of an individual creature could very well turn out *not* to have a makeup which properly accords with those norms. There is no reduction to be found here; nonetheless, such high-level intentional-stance discourse maintains its value independently of such concerns.

As we move further away from Quine's position, we find someone who takes a far less suspicious attitude towards propositional attitudes and propositional attitude

¹² Dennett, Daniel C., "Self Portrait", in S. Guttenplan (ed.), *Blackwell Companion to the Philosophy of Mind*, p. 240.

discourse. Like Ryle and Dennett, Jerry Fodor finds value in propositional attitude talk; against Quine and other eliminativists like the Churchlands, he complains that it is a ‘grotesque’ proposal that we should abandon propositional attitudes and a psychology that trades in them, because, ‘on the one hand, propositional attitude psychology works, and, on the other hand, nothing else does’.¹³ However, unlike Ryle, Fodor does not regard propositional attitude discourse as confused insofar as it allows that there are such things as minds and beliefs, and unlike Dennett, Fodor does not think the purported objects of such discourse are best characterized as convenient fictions with practical application. Although bent on finding a respectable place for propositional attitude psychology in the world of scientific endeavour, and hence a physicalist, Fodor is a realist with respect to propositional attitudes. Like Dennett and unlike Quine, Fodor’s physicalism is not reductive: he seeks to give a species of functionalist account of mental states (couched in the language of propositional attitudes). The functional specifications of each state are realized in physical objects (such as human brains), but because a particular function can be realized in multiple physical ways, those functions cannot be reduced to any particular configuration of physical matter. For Fodor, then, it would seem that psychology does not primarily seek an explanation of mental states in terms of states of physical objects; rather, it seeks an explanation of mental states in terms of functional states of an organism (broadly construed; i.e. not necessarily carbon-based). Functionalism makes plausible the idea that there is a “*level of explanation* between commonsense belief/desire psychology, on the one hand, and neurological (circuit-theoretic; generally ‘hard-science’) explanation on the other”.¹⁴ In fact, as Fodor notes,

¹³ Fodor, Jerry, “Banish DisContent”, in Jeremy Butterfield (ed.), *Language, Mind and Logic*, p.1.

¹⁴ Fodor, *A Theory of Content and Other Essays*, p. 9.

functionalism sometimes goes beyond making plausible the idea that there is an intermediate level of explanation between propositional attitude psychology and neuroscience, as it often *identifies* propositional attitudes with functional states.

Fodor's body of work represents only one species of functionalism, but it does not serve our current purposes of territorial survey to delve into the internecine struggles taking place within this arena; other issues are at hand.

With the possible exception of Ryle, those philosophers discussed so far are all physicalists on one sort or another. Some individuals are inclined to reject physicalism altogether: Chomsky, Nagel, and Williams among them, although for different reasons. Perhaps the most striking is Chomsky, who claims that we have not had a well-defined notion of the physical since Newton's time. Chomsky takes Cartesian dualism as the starting point for his reflections on physicalism. Descartes offers, according to Chomsky, a view of the physical world 'in terms of a kind of contact mechanics'.¹⁵ Certain aspects of the world do not fall under the principles that ground this view of the physical; in particular, those aspects of the world we assign as mental. This created a unification problem, which Descartes solved by positing 'a second substance whose essence is thought',¹⁶ and by attempting to show how these two substances interact (notoriously, through the pineal gland). Chomsky writes

As is well known, this programme collapsed within a generation, when Newton demonstrated that the theory of the material world was fatally inadequate, unable to account for the most elementary properties of motion. Newton had nothing to say about the ghost in the machine. He exorcised the machine, not the ghost; the Cartesian theory of mind, such as it was, remained unaffected. Newton found that bodies had unexpected ghostly properties; their 'occult quality' of action at a

¹⁵ Chomsky, Noam, "Noam Chomsky." *The Blackwell Companion to the Philosophy of Mind*, p. 156.

¹⁶ Ibid.

distance transcends the common notion of body or material object. Like many other leading scientists of the day, Newton found these results disturbing, agreeing with the Cartesians that the idea of action at a distance through vacuum is 'so great an Absurdity that I believe no Man who has in philosophical matters a competent Faculty of thinking, can ever fall into it', a reaction that is understandable, possibly even rooted in folk psychology. He concluded that we must accept that universal gravity exists, even if we cannot explain it in terms of the self-evident 'mechanical philosophy'. As many commentators have observed, this intellectual move 'set forth a new view of science', in which the goal is not to 'seek the ultimate explanations', but to find the best theoretical account we can of the phenomena of experience and experiment (Cohen, 1987)...These moves also deprive us of any determinate notion of body or matter. The world is what it is, with whatever strange properties may be discovered, including those previously called 'mental'.¹⁷

Let us first note that Chomsky's outlook appears to allow that it is legitimate for us to investigate those properties of the world which we are presently inclined to label 'mental', however mistaken that inclination may be. We may take this as an indication that Chomsky regards the term 'belief', among other mental terms, as a legitimate topic of discussion. For our present purposes, this is all we intend to extract from the quotation above regarding Chomsky's own views on propositional attitudes; but some further discussion of the points raised in that quotation will lead us to a new part of the local terrain.

Chomsky's arguments are directed at the dissolution of the mind-body problem. However, these arguments can be used to serve a different, if related, purpose. If we look at the etymology of the word *physics*, we find that one of its original meanings is nature. The proper domain of the discipline of physics is the study of nature, whatever that may be; and so it is at once an attempt to seek ultimate explanations as well as find a theoretical account of the phenomena of experience and experiment, as these phenomena are part of nature. The two goals are not exclusive. That being said, Chomsky's

¹⁷ Ibid.

observations do point to something odd about the priorities of the physicalists. If we are investigating nature, and find various aspects of it seemingly at odds with each other, why should we privilege one observed aspect or set of aspects as fundamental and assign all other aspects to ancillary roles? Suppose we have, on the one hand, a well-developed physical theory expressed in a particular terminology, and on the other, a well-developed theory of the mental expressed in different terminology. If we are attempting to unify these two disparate theories under the rubric of one or the other theoretical apparatus, we have no obvious motivation to choose either one as fundamental. One might argue that because the objects of physical study are in some way public, whereas mental entities are private, we should adopt the physical theory as fundamental. However, such an argument would have no weight with an individual possessed of solipsistic inclinations, skepticism about other minds, or any other point of view which generates philosophical worries about the proper meaning of the term 'public'. One possible strategy which avoids this issue of priority among domains of discourse could be deployed if an argument were first made to the effect that the reduction of one to the other is actually not possible. Then one might say that, since the two theories of the mental and physical cannot be directly reconciled because we cannot find a way to reduce the terminology of the one to the other, the theory which unifies them will need to be couched in a terminology alien to both, but to which both can be reduced. This is one form of the doctrine of neutral monism, variously defended by William James, Bertrand Russell, and Donald Davidson, among others. It may be helpful to look at the position of Davidson, the most recent champion of this outlook.

Davidson's monism is centered on the notion of events; in his work, physics, psychology, and other sciences are characterized as descriptive tools. We can describe certain events either as physical or mental. This approach neatly provides solutions for a number of problems, among them the problem of the reduction of mental objects, states and properties to physical objects, states and properties. Returning to the question which began this exercise, namely the question of what a belief is, the Davidsonian approach allows an answer something like the following: a belief is a kind of event, described in mentalistic terms; the same event can be redescribed in physical terms. While we could say that these points alone allow us to say that there is no motivation to reduce the mental terminology to the physical, Davidson goes further than this, denying that reduction is in fact possible, because the two descriptive terminologies have different purposes and characteristics. Among the theses he uses to argue the point are the lack of causal closure in the domain of the mental (as opposed to the causal closure constitutive of the physical domain) and the difference in purpose of descriptions in the two domains: mental descriptions serve the purpose of explanatory understanding of other minds and human behaviour, while physical descriptions serve the purpose of the formulation of laws.

Event monism coupled with descriptive pluralism has appealing and unappealing aspects. The dissolution of some problems which have plagued physicalists and dualists alike makes the position attractive. However, the ontology of events invites the introduction of a new descriptive apparatus which describes and categorizes events as events, which poses new problems. Davidson's arguments and his overall outlook have been taken to task repeatedly; we need not rehearse any of this material here. Our primary concern is with the nature of belief, if it has any nature at all. Davidson's outlook

allows one to claim that talk of beliefs is legitimate, but not in virtue of belief-talk being one way of describing events. If it were the very fact that belief-talk picks out a species of event that made such discourse legitimate, then one could develop a vocabulary of *smmental* entities, among them *reliefs* and *lesires*, and claim that these too are legitimate terms because they refer to event-types (that one would, of course, then specify). While Davidson's outlook *appears* to allow this, the legitimacy of such a *smmental* vocabulary would be questionable (by Davidson's lights) because it would serve no obvious purpose. The vocabulary of mental description, however, does serve a purpose: to explain (and predict) the behaviour of intelligent beings and describe their minds. In this sense, Davidson's attitude with respect to talk of beliefs converges to a significant degree with the attitudes of Fodor and Dennett.

It is worth making a few remarks on the attitude of Wilfrid Sellars, who, like Davidson, treats propositional attitude talk as a form of theory. Sellars' provocative essay on perception includes a story about a fictive genius named Jones.¹⁸ Having been born into a society of Ryleans, Jones develops a theory intended to explain (and perhaps predict, *à la* Dennett) the origins of overt acts by himself and his fellows. Since apparently goal-directed behaviour is not always attended by overt linguistic acts, Jones postulates inner episodes called *thoughts*, which are the primary origin of both linguistic behaviour and (most) non-verbal behaviour. Jones has great success in teaching this theory to his fellows. Their assimilation of the theory is so complete that they come to treat it as descriptive, giving articulation to real entities upon which *reports* can be made. Introduced as a practical fiction, the theory gains its own momentum, generating

¹⁸ In the development of his tale about Jones, Sellars demonstrates that the conceptualization of experience (conceived as an inner episode) penetrates to the deepest possible level, and does serious damage to the classical notion of what is given in perceptual experience.

philosophical worries in its train about such things as beliefs, impressions and sensations. From the Sellarsian perspective, the problems that this paper treats are thoroughly bound up with some version of Jones' theory; but even from that perspective, this does not count against the legitimacy of exploring those problems or the associated theory. Even if the ontology suggested by such a theory should turn out to be false, articulation of it may help in, for example, its reduction to physical theory or into an ontology of events – if that is what is desired. At the very least, discussion of these purported entities will aid us in determining the consequences of holding a theory which postulates them.

We have looked quickly at a number of the opinions of professional philosophers relating to the notion of belief, with an eye to determining what that notion consists in. The results of this cursory examination are inconclusive at best. We have an attempt to analyze beliefs as dispositions, brought by Quine and Ullian, an attempt to eliminate beliefs and other mental terms, brought by Quine, an attempt to analyze belief-talk as an heuristic tool for prediction brought by Dennett, an attempt to analyze beliefs as functional states of an organism brought by Fodor, a claim that mental properties are legitimate topics of inquiry for science brought by Chomsky, and an attempt to interpret propositional attitude discourse as only one of many legitimate forms of description brought by Davidson. Even Sellars' critique of inner episodes does not explicitly deny that we may legitimately explore the framework of the theory propounded by the mythical Jones. Many of the various ways in which talk of beliefs could be construed as legitimate or illegitimate were not discussed here, but an exhaustive survey of the literature was not the goal. We were looking for a consensus in the literature of professional philosophy to the question of what a belief is, or failing that, whether it is

reasonable to treat talk of beliefs as legitimate. Clearly, as is often the case in the philosophical literature, the jury is still out. How, then, should we proceed?

It is not clear whether a reduction of the mental to the physical is possible, let alone desirable. If it is possible, even staunch eliminative materialists like the Churchlands recognize that the vocabulary of the mental can only be eliminated if neuroscience can fully account for all of the features which we associate with mental terminology. As Paul Churchland puts it, “[i]f and when the change ever happens, it will be because we are all gleefully pulled rather than grudgingly pushed. We will be pulled, if at all, by the manifold, personal, social, and practical advantages of the new framework: by the clarity it represents, by the freedoms it makes possible, by the cruelties it diverts, and by the deeper interactions it affords.”¹⁹ Until that time arrives, we are ‘stuck’ with the old terminology. While the legitimacy of that terminology may or may not be on shaky ground, it still appears to have explanatory value for us, and we have little choice in making use of it. From the point of view of eliminativists like Quine and the Churchlands, further investigations of mental terms would lead ultimately to a list of features for which neuroscience would want to account when we purge ourselves of mental terminology altogether; from the point of view of anti-reductivists like Fodor, Dennett or Davidson, such investigations will simply add further clarity to a vocabulary already regarded as legitimate. As this is the case, further investigation of some of the features of terms in the mental vocabulary is worthwhile.

In what follows, we shall examine some features of the term ‘belief’ in the hope of adding to our understanding of its significance. In particular, we may perhaps discover

¹⁹ Churchland, P.M., "**Replies**" in *Philosophy and Phenomenological Research* Vol LVIII No.4, December 1998.

an answer to the question of what (and when) it is reasonable to believe. Before continuing, however, it is worth making some remarks upon belief which have not been covered by the cursory examination of modern notions of the term that we have just concluded, and which focus more narrowly on the issues we will raise in what follows.

We might begin by noting that each of the views sketched above presumes that a belief is to be construed as a kind of mental entity. Whether the term is regarded as empty (say, because all mental terms are empty), as poorly employed by normal users or philosophers, or as a catch-all for certain sorts of event or state, it is tacitly agreed that the term intends to pick out a mental phenomenon. As put by Lord Quinton:

Most philosophers who have in any way adverted to the nature of belief have assumed that belief is an inner state of mind, directly accessible to introspection and distinct from, though causally related to, the believer's behaviour.²⁰

The only individuals we have glanced at in the preceding discussion whose views might be construed as falling outside this fairly unobjectionable intended definition are Gilbert Ryle and W.V. Quine, considered on his own. We have seen in the brief exploration of Quine and Ullian's views that they regard, or appear to regard, belief as reducible to dispositions to behave. They do not quite go so far as to identify belief with behaviour, which allows that their views can be included under Quinton's umbrella. Quinton makes reference later in the same essay quoted above to Alexander Bain, who explicitly proposes a definition of belief in terms of behaviour. The views of Ryle and Bain aside, Quinton's proposal does seem to fit the common notion of belief. He goes on to describe the notion as analyzed by H.H. Price and R.B. Braithwaite:

²⁰ Quinton, Lord A., "Knowledge and Belief.", *Encyclopedia of Philosophy*, Paul Edwards (ed.), vols. 3-4, p. 351.

Price's mentalist definition of belief equates it with entertainment of a proposition together with assent. To entertain a proposition is to understand and attend to its meaning; when it occurs by itself, it is neutral and uncommitted as regards the proposition's truth or falsehood. Price breaks assent down into a volitional and an emotional part. He describes the volitional element as a mental act of preferring a proposition to any incompatible alternatives that have occurred to one; the emotional element is a feeling of conviction or assurance and may vary in degree. Braithwaite identifies belief in a proposition with its entertainment together with a dispositional readiness to act as if it were true. "Being ready to act as if it were true" has at first sight a suggestion of circularity, for it seems to mean being ready to act as if one believed *p*. But this can be avoided. I act as if *p* were true if I act in a way which would satisfy my desires if *p* were in fact true.²¹

There are three points in this passage which are of immediate interest. The first two concern Price's breakdown of assent. Assent is here considered to have an emotional element, viz. "a feeling of conviction or assurance". It may well be the case that all beliefs, in *human* cognitive systems, are accompanied by such a feeling, and that the feeling itself may contribute to the role that the belief plays in an individual's cognition. For example, if an individual's feeling of conviction with respect to the (propositional) content of a belief is absolute, that individual might be unable to revise his opinion in the face of contrary evidence. Certain kinds of dogmatism might be traced to just this sort of emotional aspect of assent. However, it may be unwise to consider this aspect of assent as *essential* to the understanding of the term 'belief'. It appears to be perfectly possible to imagine that there could exist a cognitive system which is capable of belief, yet incapable of emotion, and hence whose beliefs could not possess the emotional aspect of assent as described by Price.

Price's analysis of assent also provides for a volitional element, but it is not entirely clear that we should want to consider this aspect of assent as itself essential to

²¹ Ibid.

understanding the character of belief. For it to be essential, we would need also to make it a requirement of cognitive systems that are capable of belief in general that they be possessed of a *will*. Without will – unless I am mistaken - there can be no volition. It is perhaps too much of a digression to explore this particular issue here. However, we may ask the following question: on the assumption that there could exist cognitive systems capable of belief but possessed of nothing that could be identified as a will, or even just fulfilling the functional role of a will, is there anything left in the notion of assent which can separate belief from mere entertainment?

It may be possible to derive, from the basic sense of belief as a truth-regarding attitude, a notion of *commitment* towards the proposition so regarded. If this commitment is not to be described as either emotive or volitional, it must have some other description appropriate to it. For want of a better category, it may be submitted that the commitment is a commitment of reason. This is not intended to suggest that the form of commitment which appears to be implied by a truth-regarding attitude is itself necessarily *arrived at* in a rational manner; one can believe a proposition for a bad reason or set of reasons, or for no reason at all. Nonetheless, once the commitment has been made – the truth-regarding attitude adopted, in other words – its employment in cognition may be considered as constrained by various desiderata of rationality (insofar as cognition itself is constrained by such desiderata). In particular, the extent to which a cognitive system is sensitive to and seeks to avoid contradiction possesses a central role in the development of the problem which the truth-taking attitude generates. The interplay of that commitment and the allegiance of a cognitive system to some form of the law of non-contradiction

constitutes the exploration of the central problem of this paper, and so further discussion will be reserved for later chapters.

The third point of interest concerns Braithwaite's analysis, which proposes that a belief consists in the entertainment of a proposition coupled with a dispositional readiness to act as if that proposition were true. Recall that Quine and Ullian claim that a particular belief can be analyzed as being disposed to behave in certain ways given the appropriate stimulus. This claim suggests that each belief can be analyzed into a complex conjunction of conditionals, where the antecedent of each conditional is a description of a stimulus, and the consequent a description of a certain type of behaviour. Braithwaite, in contrast, employs the catch-all phrase 'be disposed to act as if it were true' without suggesting further analysis for each particular belief. Keeping Braithwaite's phrase in mind, let us reconsider the notion of an occurrent or activated belief. In the course of discussing this point of view earlier, some gestures were made in the direction of an analysis of the 'believing attitude'. In particular, it was suggested that the believing attitude could be regarded as 'taking to be true', where what is taken to be true could be a sentence, a proposition, or some other form of content. Quinton himself seems to recognize this way of construing the term, where he mentions the possibility of taking Braithwaite's analysis as circular: he sees the possibility of substituting the phrase 'as if one believed p' for the phrase 'as if p were true'. Quinton also mentions that this apparent circularity can be avoided, because acting as if p were true does not commit an individual to the truth of p; an individual who has no actual commitment to the truth of p can act in ways identical to those of an individual who is so committed. Quinton mentions it only because it is a natural move for most people who would accept the common view which he describes

above. It is a natural move to make, because it seems intuitive to say that for an individual to believe a proposition *p* is for that individual to regard *p* as true; that is, that for an individual who believes *p*, *p is true*.

This brings to the foreground a question. The position

(a) for an individual who believes *p*, *p is true*

is clearly distinct from the position

(b) for an individual who believes *p*, that individual is disposed to act as if it were true.

The second statement involves no commitment on the part of the individual to the truth of *p*. While a dispositionalist can consider the merits of (b) as a possible analysis of belief, (a) is of no real interest as it simply regenerates the problem the dispositionalist would want disposed of in the first place. For those who are partial to non-dispositional analyses, however, both (a) and (b) hold some interest. Where (a) supposes simply that the believing attitude is constituted by taking a proposition *p* to be true, (b) allows the possibility of the believing attitude being constituted by regarding a proposition *p as if* it were true, without actually committing to taking *p* to *be* true. In other words, (b) opens a possibility for an instrumentalist or pragmatic perspective with respect to belief.

Believing in the sense suggested by (b) allows that the individual concerned will license him- or herself to think and act as if *p* were true, without committing to the truth of *p*. For those who both accept Quinton's formulation of the common view and look favourably upon propositional attitude talk, this raises the question of whether (a) or (b) properly captures the sense of what it means to believe.

It is uncontroversial that the sense captured by (a) is a common view; Moore's Paradox, for example, depends upon it.²² There will be no argument made here as to whether (a) or (b) is the better formulation. It is quite possible that the term 'belief' as used by any individual shifts back and forth between the two, either with or without the individual's notice; in which case both (a) and (b) are legitimate. In what follows, the arguments made concerning the revision of belief will advert to the sense captured by (a). It may appear to the reader that the force of those arguments would not be felt if belief were construed in the manner of (b). This may only be determined by first answering the question of how deeply position (b) is intended as an analysis of belief. If the analysis penetrates to the level of the mechanics of perceptual interpretation, such that those mechanisms are understood to be operating in way identical to the way they would in a cognitive system for which position (a) is a correct description, the problem will arise notwithstanding the apparent lack of commitment. If position (b) applies more superficially, referring only to how an agent's overt behaviour may be interpreted, then position (b) *might* escape the problem. This issue will not be explored in this paper, but in traversing the territory of the problem laid out in the body of the essay, those inclined to favour position (b) may find that it aids them in understanding the options available to them.

This initial discussion began with a fanciful description of the effect felt upon first contact with the discourse of Descartes' *Meditations*. The metaphor used conveys some of the sense of this impact, and its legacy in the literature of philosophy in the modern

²² As named by Wittgenstein; the paradox explores the legitimacy of such sentences as 'Smith left the room but I don't believe it'. When belief is construed in the manner suggested by (a), and an assumption made concerning the relationship between a bare assertion and assertion of the truth of that assertion, the paradox is virtually manifest; when construed in the manner of (b), the paradox *can* be made to disappear (fancy footwork is nonetheless required).

period. It is no easy task to find a way out of the gloomy cellar to which Descartes has led us, and no one, whether by direct argument or by deflection of the philosophical issues involved, has really succeeded in doing so; at the very least, no one has ever succeeded in convincing everyone else to adopt their position. This essay cannot claim to have a hope of succeeding in this respect, where so many other arguments have failed. Rather, the aim of this paper is to raise some problems and suggest some resolutions which may be of interest to *anyone* who has wrestled with the problem of skepticism, the nature of belief, or the acquisition and revision of belief, whether they have ultimately followed the paths hacked out by Hume, Kant, Popper, Quine, Ryle, Dennett, Fodor, or anyone else. Sustaining this broad interest is certainly a lesser challenge than offering a universally acceptable answer to Cartesian skepticism, but it is challenge enough.

Core Argument

We shall begin with a statement of the central problem of this essay, and then proceed to discussion of a thought experiment. The thought experiment describes a *simple* case in which the problem is exposed.

The problem can be put succinctly in the following way. The attachment of an ascription of truth to an accepted claim results in that claim losing all future potential for rational revisability in the light of new experience: the claim becomes rationally indefeasible. If our own cognitive architecture were grounded in belief, where belief is understood as a truth-taking propositional attitude, then we would find ourselves in the position of being unable to rationally revise any of our beliefs. Patently, we rationally revise those claims that we accept *all the time*; and it is generally supposed that the proper term for describing an accepted claim is the word *belief*. So we are left with a problem.

Let us imagine a scientist. The scientist accepts a body of theories, which shapes her understanding of her work and guides her inquiries. Let us suppose that this scientist has a perceptual experience – in the course of investigating an experimental result, for example – which could be interpreted as conflicting¹ with one of the theories she accepts. How is the conflict manifested? The theory the scientist accepts yields observational consequences. One of the consequences predicted by the theory accepted contradicts the experimental result, when the phenomenal datum is interpreted in a certain way. As is

¹ The phenomenon of *recognizing* a conflict among beliefs – or more generally, any thoughts having propositional form – is left unanalyzed for the greater part of this paper. The issue is addressed at some length in the latter portion of chapter 6.

generally accepted, the scientist has two options in this situation: she can deny the correctness of the interpretation of the phenomenal datum; alternately, she can reject the theory she holds that yields a contrary observational prediction.² The rejection of this theory would constitute a change of opinion; the scientist would, in this case, revise her acceptance of the theory in question. Denying the correctness of the interpretation of the datum could take many forms: the scientist could say that the equipment was faulty, the lab assistant inept, or that some exceptional circumstance obtained and so the experimental result does not fulfil the *ceteris paribus* clause attached to the accepted theory. In any case, there are many ways (of varying degrees of plausibility) of denying that the experimental result counts as a refutation. For any particular case where such a conflict between experimental result and accepted theory occurs, there are circumstances in which the scientist will be constrained to do the one or the other.

But now let us suppose further that the scientist regards the theories that she accepts as *true*, which is to say that she *believes* them; the rest of the scenario remains the same. May the scientist, in this case, exercise either of the two options available to her as described above? The answer is no. She may exercise only her right to reject the interpretation of the datum, by means of whatever explanatory mechanism that she deems appropriate. If the scientist genuinely takes the theories that she accepts as true, then she must accept by that attribution that they describe the world, or a part of it, as it really is. Thus, when a counterexample is presented, the scientist is bound by reason alone to judge the counterexample to be only an apparent refutation of the theories she takes to be true. For how could anything that is true be refuted? It may be the case that the theories that

² Not only could she reject the theory she actually holds, but more generally, she can reject or refuse to accept any one of the entire family of possible theories which would yield that particular prediction; this consequence falls naturally out of the underdetermination of theories by evidence.

the scientist accepts as true, are *in fact* false; but it is the scientist's subjective evaluation of the theories which will drive her judgment. The actual truth or falsity of the theory or theories in question is irrelevant in this context.³

Let us imagine the same situation, but with the evaluations of theory and interpreted datum reversed; we shall now assume that, instead of accepting these theories as true, our scientist merely accepts them. One wants to ask immediately what acceptance, full stop, amounts to, as opposed to accepting as true, or taking to be true. We shall defer the answer to that question until we have made a thorough examination of the problem and some objections which arise from that examination. Now let us assume also that the scientist accepts the experimental result, interpreted in just the same way as above (i.e., conflicting with the accepted body of theory), and accepts it as *true*, or in other words *believes* it. In this case, there would be no question of doubting the reliability of the equipment, the circumstances of the recording of the datum, misinterpretations of the datum, or any other concerns that might cast doubt upon the interpreted experimental result. If the interpretation of the datum is taken as true, then doubt is an inappropriate (if even possible) attitude to take towards it. Thus if the scientist only *accepts* her theories, but takes the interpretation of the datum which conflicts with those theories as true, she is rationally constrained to reject the theories she previously accepted.

Our fictive scientist does not well represent any individual who practices science. It might be thought that, in fact, the behaviour of our scientist in the first scenario – in which she *believes* the theories which bear on her experiments – actually does correspond to the behaviour of scientists undertaking 'normal' science within a Kuhnian paradigm.

³ At least, the option to change one's mind *rationally* is not open to the scientist; for if no potentially conflicting interpretation of experience can be accepted as true and hence actually challenge her extant beliefs, then any change in opinion effected would have to be arbitrary.

Blinkered as such individuals are taken to be by the framework of the paradigm under which they operate, it might be argued, we can expect exactly this sort of scenario to develop in episodes of experiment, wherein interpretations of phenomena which do not accord with the theories which frame the paradigm are simply rejected as conflicting and ad hoc explanations for the friction are sought. However, the problem is more serious than that. In Kuhn's sociology of science, paradigm shifts occur when there arises a crisis in a scientific community, brought on by the repeated failure of experiments to yield expected data. A critical mass of such episodes is supposed, in Kuhn's view, to spark a revolution, after which the members of a scientific community toe a new party line. Perhaps Kuhn's analysis of the sociology of science is correct; we will not delay ourselves with that debate. Whether correct or not, Kuhn's model could not apply to the scientist in our scenario. For, in believing the theories which frame her conduct of the experiments she makes, there is no room for any sense of failure. No matter what phenomenal experience issues from her pokings and proddings, an interpretation of that experience may be found which will accord with the theories she already believes, and only interpretations of that sort will be admitted as beliefs. Without a sense of failure, no crisis of faith may arise which might lead to our scientist being bewitched by the siren song of some new scientific worldview.

We may put the matter in a different way. In order for our scientist to recognize that there may be a problem with one of her extant beliefs – in the present case, a particular theory that, coupled with some initial conditions, yields a prediction about the outcome of her experiment – she must actually come to possess an interpreted perceptual belief that is at variance with her expectation. It is reasonable to suppose that among the

minimal desiderata of rational thought, there should be included respect for an epistemic version of the law of non-contradiction; one may not believe both A and $\sim A$ at the same time. Yet it seems that precisely such a cognitive condition must be achieved, at least momentarily, simply in order to recognize the friction between two beliefs. Such recognition could then occasion the engagement of mental processes which are equipped to deal with such friction. Our problem arises, within a belief-based cognitive system, because there is a case to be made regarding the impossibility of achieving such a cognitive condition at all. If a cognitive system is so constructed that there is no way in which it may come to possess a belief (perceptual or otherwise) that conflicts with any of the extant beliefs of that system, no friction will be recognized and no crisis precipitated which would require resolution. Without any rational motivation to reexamine and revise, there can be no rational change of opinion. We are exploring here the idea that a cognitive system whose processes are grounded in beliefs faces this difficulty.

In order to further illuminate the issue, we would do well to discuss two matters which arise for any cognitive agent: the epistemic evaluation of ideas, and the order of events which leads to the acceptance and evaluation of those ideas.

The epistemic evaluation of ideas, thoughts, or episodes of thinking possessing propositional form, consists in the assignment – by whatever method – of a particular kind of ‘score’ to said ideas. There are many proposals on offer for describing the metric which best captures our actual practices. Minimally, there are two values which may be ascribed to any particular idea: either that it is true, or that it is false. Some systems attempt to offer gradations between the two endpoints designated by ‘true’ and ‘false’; for example, ‘improbable’, ‘probable’, ‘best explanation’, and so on. It is generally

agreed that the highest (most positive) evaluation that may be ascribed to an idea is the absolute value 'true'. However we manage to parse the lesser evaluations that may be made, an ascription of truth made of a particular idea bestows upon that idea (for the subject ascribing the value) a power akin to the power of a trump card in a game of bridge. The analogy is by no means perfect; in a bridge game, the cards in the trump suit are themselves graded and can be used against each other in a particular hand, and the outcome of any hand is a zero-sum game: there is always a single card which defeats all others. The analogy would be more exact if there existed a variation of bridge in which all the cards of the trump suit possessed equal powers with respect to each other, trumping only the cards of the lesser suits. In other respects, however, the analogy is apt.

When confronted with a situation in which there is, on the one hand, a *belief* concerning either some theory or some observational matter of fact, and on the other, an interpretation of some phenomenal experience which offers the potential for conflict with this belief, the belief will trump the potentially conflicting interpretation, forcing the rejection of that interpretation even before it may be considered an actual challenge to said belief. The believing attitude is a truth-ascribing attitude. To make such an ascription of a proposition is to make a claim that the proposition represents, is indicative of, or is in some way tied to the way things actually *are*; a claim about the way the world *is*. From the point of view of a believing subject, the potentially conflicting interpretation of a particular experience on offer possesses no such ascription; it is, unless further discussion clarifies the matter one way or another, an open question as to whether any epistemic evaluation has been made of it at all. It may not even be correct to say of this potentially conflicting interpretation that it is, in the eyes of the believing subject, even so much as a

genuine possible description of some feature of the world it considers itself to *actually* inhabit. Before completing our discussion of evaluative inequality, we should discuss these matters, as they arise, within the context of discussion of the order of events by which a potential interpretation of phenomenal experience may come to be accepted by a believing subject.

There is probably no particular consensus to be found among philosophers as to what exact sequence of events accurately describes the transformation of a raw phenomenal apprehension into a conceptually specified perceptual experience. Our discussion does not require that we go into a great deal of detail regarding such a sequence, and so allows us to avoid a significant proportion of the debate to be found here. Nonetheless, even the bare specification that will be discussed below will probably itself be found contestable.

In particular, the sequencing required for the generation of the problem we are discussing itself requires an assumption concerning the boundary between mind and world. The framework for our problem assumes that there *is* such a boundary; the primary consequence of such an assumption is that all perception is potentially subject to skeptical worries of one sort or another. The conscious mind never perceives anything beyond the mind-world boundary in an unmediated way. At the boundary itself – wherever it may lie – the conscious mind may apprehend a non-conceptualized *content*, in the sense that it may become aware that something is happening. Certainly we may allow that, in a belief-based system, this awareness that something is happening may immediately become a *belief* to the effect that something is happening. However, it is difficult to imagine that such a belief could conflict with any other beliefs the subject

may already hold; hence, the sequence of events that may allow the formation of such a belief is of little interest to us. Our central concern must be with the formation of beliefs with conceptualized content. At least, our concern must be with beliefs with more *detailed* conceptualized content; for it may be allowed that in the act of becoming aware of *something* happening, the something is at least minimally conceptualized as *something*. But by our assumption of the existence of a mind-world boundary, the *something* in question would not be any sort of Kantian noumenon, and so not a direct perception of anything beyond the boundary; this *something* would be only a minimally specified phenomenal object of perception.

Whether or not such minimal perceptual beliefs of the sort just described are possible, most perceptual beliefs are greatly enriched by the application of various concepts. It is probably not unfair to say that the process by which untreated phenomenal experience comes to be so enriched is generally envisaged by way of a conveyor-belt metaphor: the untreated experience rolls along the conveyor, and during its travels is shaped and packaged by a variety of mechanisms. Such a metaphor does not entail the notorious Cartesian Theatre model of perception; as in Dennett's multiple-drafts model, we may envisage the perceptual item being hauled off the conveyor at any time during its voyage, or more radically, that we consciously experience (at least some of) the various stages of the process of interpretation.⁴ We are not constrained to imagine that a potential percept must be completely processed before being delivered to a conscious homunculus waiting at the retail-outlet attached to the perceptual manufactory. The metaphor does not favour one or the other model, and nothing depends upon our choosing either one for the

⁴ Although we shall not explore Dennett's model, it may be mentioned that under his conception we might imagine that there are multiple conveyor-belts operating simultaneously on the perceptual factory floor. The problem we are exploring might occur on any one of those conveyors.

purposes of exposition of the central problem. Whether or not the item becomes a consciously apprehended phenomenal object of perception as soon as it first rolls onto the conveyor, or better, that ‘rolling onto the conveyor’ is a metaphor for becoming conscious of the object as *some thing*, we may usefully describe the action of the mechanisms encountered along the conveyor as fixing an interpretation upon that phenomenal object.

The action of the mechanisms upon that object yield an interpretation as to what *sort* of thing is happening. However, while the phenomenal object may already be considered the object of a perceptual belief at this initial point, we must ask whether it becomes a more definite – that is, a more specific – belief as it passes through each stage of its processing, or whether the interpretation gradually being pressed upon it remains itself unbelieved until some later stage of inspection. Suppose that we are dealing with the raw material of a visual perceptual belief. We may say that immediately after the initial stage of becoming a belief that something is happening, a conceptual framework is applied that produces a belief that the something that is happening is a *visual* something. Following this, a variety of concepts of increasing complexity may be applied, beginning, perhaps, with concepts relating to space; the subject’s perceptual processor may interpret variations in the pattern, shading and coloring of the visual phenomenon as a particular shape against a background, determining that a certain variation in colour or shading represents an edge rather than just a change in surface pattern, that the foreground shape has a particular size, and so on. This portion of the process of interpretation of raw phenomenal experience is commonly referred to as “bottom-up” interpretation.⁵ Perhaps

⁵ Earlier, footnote reference was made to the Sellarsian critique of givenness and the deep level to which concepts reach in the process of perception. While this paper does not treat of the problem of knowledge,

the product of this particular layer of interpretation is itself a belief, rather than an interpretation awaiting later inspection. Further layers of interpretation which may be imposed upon the phenomenon may produce a ‘final’ interpretation, in which all those aspects of the phenomenon – that is, all those aspects considered relevant to the context in which the perceptual episode is taking place – that may be open to interpretation have been exhausted. In the particular case we are examining here, let us suppose that the final interpretation yields a visual experience of a cup of tea sitting on a desk. Is this final interpretation produced as a fully-formed belief, or must it first be submitted to an inspection process before being deemed acceptable?

Although it is of value for us to know the answer to this question, it appears that the problem will remain undissolved no matter which answer we choose. Suppose that each stage of the interpretative process yields not merely an enriched interpretation, but a belief. What mechanisms might we posit as being responsible for the development of such interpretations? In the earliest, “bottom-up” stages, it is not implausible to suggest that the processes controlling the interpretation are innate;⁶ hence the family of

clearly some sort of immediate apprehension is implied in the picture of the process of perception being described here. While what is produced by ‘bottom up’ processing may be taken immediately by the perceiving subject as a belief (a taking-to-be-true), this immediate taking is not apprehension of some non-conceptualized given, as is made clear. (Even the very earliest possible stage of perceptual apprehension where what is apprehended is merely ‘something’, that apprehension is still conceptualized as *something*.) But the concepts upon which such a ‘bottom-up’ belief is built, it is suggested here, are themselves unalterable for human perceivers. Although these innate concepts used in bottom-up processing may participate in a conceptual-holist framework as building-blocks or elements in other complex concepts acquired through development, these ‘basic’ concepts are themselves inviolate. This is not to suggest that they are not incorrect with respect to reality. Quantum physics militates against the carving up of the physical world into sharply delineated objects; even if quantum physics turns out to be correct, human perceivers would be unable to perceive a ‘quantum reality’ without significant alteration of both our sensory organs and the innate concepts used in the bottom-up processing on the raw data of sensation. In this way (among others, perhaps) the collection of concepts used in bottom-up processing have none of the revisability or semantic plasticity which concepts acquired in the course of our development clearly possess.

⁶ Insofar as we are interested in parsing the process of *human* perception. No concept is *necessarily* innate, and we may imagine cognitive systems which generate all of their concepts – even those used in bottom-up processing – from some internal generative resources. It is implausible to think that human cognitive

‘primitive’ concepts which are used to determine such low-level features as foreground and background, edge and surface within a visual phenomenon may well follow a fixed procedure which itself produces a belief. The higher-order interpretative processes, however, those which in our example yield the interpretation of ‘a cup of tea sitting on a desk’, must draw their power from conceptual resources which are learned. These later stages are often referred to as “top-down” interpretation. It is with respect to this portion of the process of perception that it is sometimes supposed that a thinking subject’s extant beliefs may come to influence that process.

If this train of thought is unobjectionable, we develop the problem in the following way. If the higher-order interpretative processes are executed according to the constraints imposed by an extant set of beliefs, then we may expect that the subject will only be able to interpret the phenomena of experience in ways which accord with those beliefs. This potential description of the process has at least one positive aspect: because the interpretations are molded in such a way as to respect the constraints of an extant set of beliefs, no future inspection stage need be posited; for such an inspection would only serve the purpose of ensuring that the interpretation offered did not have the potential for conflicting with any extant beliefs. The process as described here already ensures such a condition. If this description of the order of events were correct, then, it would be a relatively efficient process. However, constraining the process of interpretation in precisely this way leads us immediately to the central problem discussed earlier. This

systems are so resourceful; but on the assumption that they are so resourceful, we would need to maintain that those concepts which become incorporated into ‘bottom-up’ processing mechanisms remain inviolate in the overall conceptual architecture of the system. We cannot assume that non-human cognitive systems must be similarly constrained; for certain such systems, there may be no division into bottom-up- and top-down-processing-specific sets of concepts. Where innateness is mentioned in this paper, it should not be taken as an endorsement of the necessity of innate concepts, structures, or anything else. Positive references made to innateness here are always guided by practical considerations of what it may be plausible to attribute to human cognitive systems.

method of processing leaves no possibility for the cognitive system of our subject to produce interpretations which may conflict with her extant beliefs, and so prevents the possibility of any of those extant beliefs ever coming under any sort of challenge. As a result, such a cognitive agent would never have any rational motivation to question any of those extant beliefs and open them to the possibility of revision.⁷

From the foregoing, we may conclude that a process which yields a new, more complex belief at each stage of its endeavours - constrained as it is by the extant set of the subject's beliefs - is an unsatisfactory candidate for a correct description of our own methods of perception. The alternative on offer is that an interpretation is not granted the status of a belief until after passing through an inspection process. This alternative may not fare any better.

⁷ Although this proposal may seem implausible, a sophisticated form of it is being developed as a live research programme in artificial intelligence circles, but with an interesting twist. A paper by Volker Stephan and Horst-Michael Gross entitled "*Neural Anticipative Architecture For Expectation Driven Perception*", presented at the 2001 Systems, Man, and Cybernetics Conference, explores the construction of a robot which interprets visual information based on expectations as it negotiates a simple maze. Stephan and Gross' robot has stored a set of statistical regularities obtaining between the operation of its motors and optical flow patterns, obtained from prior traversal of a different static maze. On the basis of these 'known' regularities and a set of current optical flow data, the robot may compute a set of 'hypotheses' about expected optical flow, with each possibility for expected optical flow assigned a different confidence score. After such computation, the next actual optical flow obtained from the robot's camera is also assigned a confidence score. All of these data are then fused through calculation and the highest-scoring result of that fusion is accepted as the correct interpretation. What is interesting about this approach in the context of our discussion is that it suggests an alternative mechanism for modification of extant beliefs; for, at the end of the process none of the original interpretations from the expectation-level or the real optical flow-level survive; only the compromise survives to become the basis for forming an expectation for the next cycle of calculations. Yet insofar as we are interested in human perception, the Stephan-Gross robot is a poor model. The phenomenal experience a human cognitive agent would apprehend as a result of the interpretation generated as acceptable by the Stephan-Gross fusion-calculation is not consonant with what our real experience in such situations as those modeled by Stephan and Gross would be. Suppose, for example, that the robot turns a corner and enters a corridor. The robot forms an expectation for future optical flow based on current optical flow. Now suppose that, after this expectation has been formed, a wall is inserted in the middle of the corridor, blocking passage through it. Our own experience in such a situation would be an immediate apprehension of a wall, notwithstanding our earlier expectations about the corridor. The robot, on the other hand, would 'experience' a compromised perception which is neither wall nor open corridor; or worse, given that the robot would constantly modify its expectation through repeated resampling of optical flow and subsequent fusion-calculation, would 'experience' the wall *slowly coming into being*. Neither 'experience' would be a good fit with human perceptual experience of the same situation.

This second proposed description of the process of perception may be subdivided into two. The first falls prey to the same difficulties as the proposal just discussed above, but a brief discussion will help illuminate the second. In this first version, the unprocessed phenomenon of a particular experience is subjected to interpretative processes similar to those described in the proposal earlier discussed. These processes are constrained by the extant beliefs of our subject, and so, as with the earlier proposal, the only interpretations which may be made are those which will accord with those extant beliefs. In this case, the passage of the interpreted product through the inspection phase will be no more than a rubber-stamp approval; for the inspection itself must consist in a check of the interpretation for consistency with the extant set of beliefs.

The first variant of the second proposed description, then, seems no great advance upon the first proposal, which did not require any separate inspection of interpretations prior to their approval. Clearly the problem lies with the suggested constraints which have been placed upon the processes of interpretation themselves. If the interpretative mechanisms used in the processing of perceptions are *not* to be constrained by a subject's extant beliefs, how else may they be fashioned? A somewhat less restrictive harness may be proposed: rather than being constrained by what is believed, the mechanisms may only be limited by what the subject can *conceive*. Whatever collection of conceptual resources a subject may have developed may contribute to the process of interpretation. In this way, the process of interpretation will not be *bound* to produce candidates for acceptance as fully-processed perceptions which accord with how the subject actually takes the world to be.

Unfortunately, the situation is not much improved by the deployment of this tactic; the candidates produced must still be approved by the inspection process before being admitted as perceptual beliefs or judgments. The inspection process itself is constrained by the subject's extant beliefs, and any candidate interpretation which has the potential to conflict with any of the extant beliefs with which it is compared by the inspection process will be summarily rejected by that process before it has any possibility of being accepted. The extant beliefs will act as a trump card, to draw on our earlier analogy, against these potential interpretations; there is never any question of these potential interpretations posing an actual challenge to any of the extant beliefs.

We have examined three candidates for a minimal description of the process by which raw phenomenal experience is transformed into fully fledged belief. We began with a description which supposed that the interpretative mechanisms are constrained by extant beliefs, and that those mechanisms themselves produce, at each stage of refinement, a fully-formed belief. We then looked over a pair of alternate descriptions, both of which employed the device of deferring epistemic evaluation until the completion of an inspection-stage. One of these assumed the same constraints on the mechanisms of interpretation as the first description explored; the other loosened those constraints by suggesting that the interpretative mechanisms need abide only by what the subject may conceive, rather than what they actually believe. It may have occurred to the reader that there is an alternative as yet left unmentioned.

Let us suppose that the mechanisms of interpretation are loosely constrained, that is, by what is conceivable, in the manner of the third option discussed above, but that the product of these mechanisms is a fully-formed belief, in the manner of the first option.

There is no deferred evaluation, and no separate mechanism of inspection which performs that evaluation. On such a proposal, we might suspect that there is a way to defeat the problem as posed above. If a particular product of the interpretative mechanism conflicted with one or another of the extant beliefs of the subject, that interpretation would constitute a genuine challenge to the beliefs concerned. By achieving this cognitive condition of genuine challenge, or genuine conflict, our subject would be in a position to recognize that something was amiss. The problem we have posed as the central concern of this paper would be resolved, then, by the adoption of this fourth schematic description of the processing of phenomenal experience. However, this final proposal as described is underspecified in crucial ways; further specification may reveal that the problem is not so easily laid to rest as the proposal suggests.

If the products of our interpretative processor were indeed bound only by what the subject may conceive, we must ask how it is that such a mechanism manages to arrive at some particular interpretation rather than another. We have been discussing the actions of the interpretative mechanism on the linear metaphor of a conveyor-belt in a factory. Raw phenomenal experience, produced at the boundary of mind and world, is fed into the interpretative mechanism and duly processed. In keeping with the metaphor offered, we imagine that the final product of these labours is a *single* interpretation. If the interpretative mechanism is capable of producing different interpretations of this one packet of information, what is it that constrains this mechanism to produce one interpretation rather than another? There is only one plausible source, within the confines of this linear metaphor, for such constraints: the subject's extant beliefs. By drawing on the current state of the set of beliefs which our subject holds, the interpreter may be set

into a 'default' configuration which leads it to process the information passing through it in a particular way. However, it is relatively easy to conclude that such a constraint would produce only interpretations which accorded with those extant beliefs, and so this solution would be no advance on the very first proposal we discussed earlier.

Let us suppose instead, then, that the initial setting of the mechanism is completely unbound: information which passes through it for processing is subjected to a totally random configuration, which may produce any conceivable interpretation. Since the product of the mechanism is here considered to be a belief, we may expect some rather odd experiential consequences which do not well accord with the felt regularity of our experience. As I am writing this, from moment to moment, I find I am experiencing a relatively clear and definite apprehension of a cup of tea on the desk next to my keyboard. Yet the information which provides the raw material for this interpretation of my local environment may certainly have been interpreted in any number of ways. If the interpretative mechanism were truly random, and its products were fully-formed beliefs, I might expect that from moment to moment, my apprehension of my local environment may undergo a radical shift. Now I am perceiving a cup of tea; now, a cup of hemlock blended with beige paint; now, an oddly-shaped silent visitor from an alien planet; now, a peculiar monument erected upon a vast wooden plain which I look down upon from the perspective of a giant.⁸ Since such alterations of interpretation are not a regular feature of our experience, this variant proposal seems a rather implausible candidate for a description of the interpretative mechanism we are attempting to specify.

One final alternative presents itself. If we at least partially abandon the linear metaphor of the conveyor, we may allow that the interpretative mechanism produces not

⁸ Apologies are offered for my limited imagination.

one interpretation during any single iteration of its operations, but *many*. Then for any single packet of raw information submitted to the processor, it will produce a group of interpretations – though perhaps not all possible interpretations, which even for a subject of limited conceptual resources may be infinite – all of which would be beliefs. Clearly, though, implementation of such a process would yield extremely odd experiential effects. A thinking subject, so described, will at any particular moment be bombarded with simultaneous apprehension of a variety of perceptual beliefs, superimposed upon each other. The cup of tea would be at once a cup of tinctured hemlock, an alien visitor, and a bizarre monument dominating a miniature world. As with the earlier variant, this too must be discarded as entailing subjective experience far outside what may be considered normal.

Unless, that is, we were to allow that the evaluation of these multiple interpretations were deferred until they had been passed to an inspection-process, which would choose among them; but this suggestion confronts us immediately with the problem of the criterion such an inspector may use to make its selection, and as may now be suspected, the only candidate on offer for such a criterion is consistency with the subject's extant beliefs. If the evaluation of each interpretation is constrained in this way, we may expect that the only interpretations which may be allowed to pass would be those which do not conflict with those extant beliefs, and so the possibility of a challenge is again eliminated. Our problem remains.

As mentioned above, the problem as described hinges on two points: first, an evaluative inequality between what is already believed, and any new information which enters the cognitive system; second, the assumption of a particular sequence of events by

which new information is processed. A qualification should be expressed; in exploring proposals for describing that sequence, we discovered that in some cases no evaluative inequality arises at all. Information is processed, with the output appearing as a fully-formed belief. It just so happens that in each of the proposals which followed this route, the acceptable products of the interpretative mechanisms described turn out to be beliefs that accord with the subject's extant beliefs. Not even a *potential* conflict may ever arise; no potentially conflicting interpretations can be made. It seemed only reasonable to discuss such sequences, as they are *prima facie* feasible proposals for the production of perceptions; however, it must be admitted that they are relatively implausible. Those proposals which include deferred evaluation and an inspection-stage of some sort seem more plausible; and in these proposals, as we shall now discuss, there seems to be some room for the generation of potentially conflicting interpretations to take place. Therefore the problem plausibly stems not only from the assumption of a general scheme for the sequence of events that enter into the process of perception, but also the epistemic evaluative inequality between potential percepts and established beliefs.

What sort of evaluative inequalities do the inspection-stage proposals introduce? The only epistemic valuation which is appropriate to these interpretations of experience, in advance of any sort of inspection, is that of being *unknown*. Beyond this, the inspection mechanism may employ different systems of measurement to arrive at more considered evaluations of each interpretation.

In the foregoing, we have assumed that such a mechanism would employ only the simplest system possible: one which assigns values of 'true', 'false', or 'unknown', based on whatever criteria are appropriate to its labours. Yet, because those criteria are bound to

the subject's extant beliefs, the situation is not an improvement over earlier proposals, even if further gradations of evaluation are admitted. Let us take the best possible case. We assume that the interpretative mechanisms are capable of generating interpretations of the raw phenomena which are not necessarily compatible with the subject's extant beliefs. We further assume that the process of generating such interpretations is not entirely random, restricting the products to what is, say, not merely conceivable but also plausible. We might argue that such a restriction may be implemented by suggesting that the mechanisms of interpretation may perform their operations on the basis that, in generating each interpretation, these mechanisms must use the subject's extant beliefs to inform their operations but may also reserve a small subset of the conceptual resources, upon which these beliefs are fixed, to treat as variables; values being assigned to those variables at random. This restriction reduces the number of interpretations which our mechanism may produce to a manageable level, and so makes this present proposal more plausibly descriptive of our own processes of interpretation. Let us then assume that these interpretations are fed, either serially or in parallel, to the inspection mechanism. The inspection mechanism must now epistemically evaluate these possible interpretations by comparing them with what is taken as true by our subject: in other words, the subject's extant beliefs.

We may now apply this framework to a concrete example; let us imagine that one of the potential interpretations submitted for review can be described as seeing, at time t , a cup of hemlock blended with beige paint upon a desk. The inspection process will duly compare this interpretation with the extant beliefs which the subject holds. We may reasonably expect that this interpretation would be rejected for conflicting with those

extant beliefs; we might find that the subject believes that, at time $t-1$, they perceived a cup of tea on the desk, and furthermore believes that cups of tea do not simply disappear only to be instantaneously replaced with identical cups filled with other substances. This pair of beliefs, coupled with some plausible *ceteris paribus* clauses, can be regarded as a believed theory about the subject's local environment, and has the consequence that at time t there ought to be a cup of tea perceivable on the desk. Hence the interpretation on offer will be rejected as false, as will any interpretation which does not accord with these extant beliefs. Despite the fact that interpretation of the visual experience in question as a cup of hemlock is just as permissible an interpretation of the experience as a cup of tea, the interpretation cannot challenge any extant beliefs and force the subject to reconsider them, as it has an epistemic status of unknown.

It seems that any attempt - on the assumptions discussed above - to admit as a belief an interpretation of experience that potentially conflicts with extant beliefs will fail. Perhaps, instead of depending upon the absolute judgments of truth or falsity which we have been assuming as available evaluations, we might try an alternate tack. For example, we might introduce *probability* as an evaluative scale for our subject's epistemic judgments. Instead of treating belief as a truth-ascribing attitude, we might say instead that belief is a matter of degree. Insofar as one may ascribe a probability to a particular theory one holds, any consequences derived from that theory will likewise possess probability values parasitic on the value assigned to the theory.

To illustrate, let us return to our scientist. Suppose there is a theory our scientist holds subjectively probable to some degree. This theory entails, in conjunction with suitable initial conditions, the occurrence of an observable phenomenon A. The degree

of confidence that our scientist has in the observable outcome A should be parasitic on the degree of confidence with which she holds the original theory and the initial conditions which are used to generate the expectation of the occurrence of A; let us suppose that the computation yields a probability assignment of .75 to A. That leaves a remaining confidence interval of .25 which may be assigned to the set of results which may occur and which are *not* A; in other words, our scientist should have a degree of confidence of .25 in the occurrence of $\sim A$.⁹ Now suppose our scientist conducts an experiment and obtains an observational result. The observable phenomena must now be interpreted. The only conceivable interpretations (from our scientist's perspective) available for the phenomena are either A or $\sim A$. If our scientist's prior probabilized belief in her theory *guides* the workings of the interpretative mechanism, then we might surmise that the mechanism would be constrained to produce an interpretation of the observed phenomenon as A rather than any member of the set designated by $\sim A$. If the mechanism were not so constrained, it might produce all available interpretations and submit them to an inspection process. The inspector may be expected to compute that an interpretation of the phenomenon as A has the higher probability, when taking into account our scientist's prior (probabilized) beliefs concerning what results may be expected; hence the inspection process would grant passage only to an interpretation of the phenomenon as an instance of A. In such circumstances, we may expect no potential for challenge to our scientist's extant beliefs concerning her theory or its expected observational consequences. Even if the experiment were repeated many times, with the same phenomenal results observed on each execution, no challenge to her existing beliefs need

⁹ A and $\sim A$, of course, are mutually exclusive and exhaustive of the possible outcomes for the situation in question.

be recognized. For despite the fact that repeated experiment and interpretation in this scenario would issue only in interpretations of the phenomena as being of type A, our scientist may simply allow that more experiment is needed in order to produce the result $\sim A$ which is expected to appear in 25% of all trials.

A subjective Bayesian might object that the introduction of new information, in the form of approved interpretations of the outcome of our scientist's experiments, ought to alter the original probability assignment that said scientist holds with respect to her theory and its consequences. If we return to the case discussed above, we may see that such conditionalization of our scientist's extant beliefs based on new information produces undesirable effects. Consider that the observed outcome of the experiment may be interpreted as either A or one of the members of the set named by $\sim A$. In this case, because the interpretation and inspection functions will always choose A over $\sim A$, we would observe a positive feedback effect. With repeated experimental trials, the probability of the theory relative to the evidence will increase asymptotically towards a value of 1. Certainly this constitutes an alteration of her belief in the theory under examination, in the sense that her confidence in that theory has changed; but this is not the sort of responsible, rational change in belief which the problem we are discussing is concerned with. There is no possibility here for our scientist to come to reject the theory, or merely reduce her degree of confidence in it, through the accumulation of evidential episodes which are interpreted as $\sim A$. If a probabilistic approach to interpretation and perception cannot effect an actual rejection, or at the very least a reconsideration of a prior held belief, then this approach does not solve our difficulty.

It is worth pointing out that, under a probabilistic understanding of the process of interpretation in perception, the portion of perceptual interpretation delivered by bottom-up processes prior to the top-down application of learned concepts may be regarded as having a probability of 1. That is, just as we earlier allowed that such primary stage, bottom-up processing may issue in a belief (that foreground is so distinguished from background, that such and such colours are present, and so on), so we may also grant to the probabilist that this portion of a perception has the highest possible initial probability. However, the high probability assignment granted does not contribute to the probability assignment that may be made on a completed perception which is built upon such initial conceptualization of raw phenomenal data. This initial stage of processing provides a scaffold upon which more refined interpretation may be hung; however, such scaffolding accommodates a variety of such refined interpretations, each mutually exclusive with respect to one another. For this reason, the high probability of the results of this stage cannot be regarded as contributing to the plausibility of any one of the interpretations that may be built upon it; or, if we grant that it does, its influence upon the final probability assignment must be insignificant. If it were not insignificant, it would have the effect of rendering all of the mutually incompatible interpretations which could be constructed upon this primitive conceptualization simultaneously plausible. Hence the outcomes of the two cases described immediately above are unchanged by the clarification noted in this paragraph. In the first case, we are concerned with the influence of prior (probabilized) belief upon that portion of an episode of perceptual interpretation which may be characterized as top-down processing; in the second case, we are concerned with the influence of prior (probabilized) beliefs on the deliberations of an inspection

mechanism deciding between two or more top-down interpretations. In each case, that portion of the perceptual interpretation delivered by bottom-up processing passes into the completed perception, and is therefore clearly not part of what is being constructed (first case) or decided upon (second case). Although the completed perception is experienced as seamless whole, the status of that portion provided by bottom-up processing (as either believed or considered as possessing a probability assignment of 1) is quite separate from the status of the top-down interpretations produced. This point may have been evident from earlier discussion, and with suitable modification is certainly applicable to our examination of non-probabilistic contexts of interpretation in perception; but it seems prudent to emphasize it here. The issue of how assessment of interpretation proceeds in the face of this division between bottom-up and top-down processing will be examined in more detail, in a set of 'case studies' discussed in the chapter on fallibilism.

It must be admitted that to attempt the construction of a probabilistic model of the interpretation of perceptual experience is to tread on thin ice. To my knowledge, no academic Bayesians are recommending such an approach; and classical Bayesian decision theory begins after such perceptual interpretation is finished: that discipline works from whole cloth, rather than unprocessed flax. Some technical papers have been produced which attempt to implement probabilism in perception,¹⁰ but the models proposed seem to fall afoul of the problem we are discussing. In those papers examined, no strategies are discussed for avoiding it.

¹⁰ See, for example: Mamassian, Pascal, Landy, Michael, and Maloney, Laurence T., *Bayesian Modelling of Visual Perception*, appearing in R. P. N. Rao, B. A. Olshausen & M. S. Lewicki (Eds.), *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13-36). Cambridge, MA: MIT Press. No pretense is being made here to the effect that an exhaustive study of approaches in this area has been conducted.

Probabilism in some form may very well have a role to play in the epistemic evaluation of perceptions, once those perceptions have been formed and already accepted. However, probabilism does not seem to be of much use in pursuit of the dissolution of our central problem.

Our discussion has attempted to show that, from a collection of relatively common assumptions concerning the relationship between perception, interpretation and belief, a quite unlikely result may be derived. Not just that change of belief is difficult, but that it may be impossible. Either the logic employed in the foregoing arguments is faulty, or there remains some unexplored variety of the family of processes discussed which reveals the problem to be no problem at all. I think that there is a variation which we have not yet discussed, which does *dissolve* the problem; yet it is not one which shows there to have been no problem, as it does not depend upon the truth-taking attitude which characterizes belief.

In describing the initial thought-experiment involving our scientist, reference was made to an undefined notion of *acceptance*. In that part of our discussion, the term was employed mainly to intimate the role that evaluative inequality plays in the generation of the problem. A believed theory was paired with a conflicting accepted interpretation of experience, with the result that the interpretation was rejected; and an accepted theory was paired with a conflicting believed interpretation, with the result that the theory was rejected. We have seen that achieving a cognitive condition of having simultaneously a prior belief concerning some aspect of the world and a conflicting interpreted perceptual belief seems to be a difficult proposal, given the influence that prior beliefs have on both interpretation and judgment. A notion of acceptance may have a role to play in escaping

this problem. The cognitive condition wanted for the recognition of conflict, which is needed in order to motivate a reevaluation, does not require that the two parties to the conflict must be *beliefs*. It requires only that there be an evaluative equality between the two parties. It will be proposed, in a later chapter, that we adopt a theory of cognition founded on acceptance as opposed to belief, in hopes that this alternative characterization disposes of the problem discussed here by allowing the generation of such an equality. That proposal may fail, but the exploration is worth undertaking.

However, we should not step directly into that discussion; the bare statement of the problem made here does not touch upon some important questions that may be raised in defense of a theory of cognition grounded in belief. We shall now turn to the exploration of some of these issues. As new objections are raised, there will be occasions on which parts of the discussion above will bear repeating in a different key; afterwards, we shall try singing a new theme altogether.

Cognitive Systems: Ideals and Limitations

An objection may be made that taking a claim to be true does not create the global problem of belief revision that I have so far suggested that it does. I have described a thought-experiment in which a scientist is supposed to be testing a theory which she already believes. Generally speaking, it may be objected, scientists do not test theories which they already believe; even more generally, people do not take the trouble to verify anything that they already believe. Conflict - or apparent conflict – usually arises in the course of investigating some problem a few steps removed from the immediate domain of the original theory, where an observational result is obtained and only later is it noticed that the result conflicts with what the original theory predicts would obtain in the same situation. In such a situation, the scientist could come to believe both the original theory - and by extension, its consequences – as well as the theory whose observational consequences conflict with the original. Her later recognition of the conflict between the two may indeed produce a mental schism, but this can be resolved by reinvestigation of both theories with an eye to rejection or adjustment of the one or the other. The mental schism would take the form of a temporary rejection of the truth of both theories, pending such a reinvestigation.

It is agreed that, if it were possible to come to believe two theories or opinions that conflicted or whose logical consequences conflicted, then recognition of that conflict would produce a temporary suspension of belief and a drive to reëvaluate. The first

question to answer is whether such a situation is possible. This question can only be answered by determining to what degree a certain type of mental disunity or cognitive compartmentalization is possible.

A. Compartmentalizing Cognition I

As a first pass at evaluating this possibility, Let us make use of the terminological device of a ‘mental context’. We won’t go into a discussion, here, of how the phrase ‘mental context’ ought to be delineated in general, and this ‘first pass’ is not intended as a serious option for consideration; rather, the idea here is to discuss some of the sorts of issues which *might* arise if we were to consider non-holistic theories of cognition as offering a solution to the central problem of the paper. Afterwards, we will look more closely at a particular ‘compartmentalizing’ framework offered by Christopher Cherniak which shows particular promise.

Let us assume, then, that our scientist inhabits a separate mental context each time she investigates a different theory. The problem can then be approached as a question of how sensitive mental contexts are to the contents of other mental contexts, or more generally a question of what should be included in a particular context. What should be considered as part of the constitutive framework of each context? If our scientist is investigating the refractive properties of a particular kind of physical surface, what would the mental context of her investigation need to include? She would probably *not* require an understanding of the structure, from an engineering standpoint, of the equipment she is using to conduct her experiments; the mental context should include, however, an

understanding of various concepts and theories which contribute to her understanding of the problem under investigation, including concepts of light, refraction, physical objects, and so on. Most of these desiderata would inhabit the background, so to speak, of the context of the investigation, and would not be the focus of the scientist's attention.

In order for the objection described above to make sense, we must assume that the mental context of the scientist's investigation does not include any part of a theory which she already believes, *and for which she has already deduced observational consequences she would regard as conflicting with a result that she would obtain in the current investigation.* If such a theory previously accepted as being true were to play a role in the mental context she inhabits while investigating the refractive properties of the surface, we would be faced with the problem as described earlier in this essay. It is not clear to me whether this sort of contextualization of thought is genuinely possible. If mental contexts were to be delineated by the focus of a scientist's attention, where all concepts and beliefs and the other mental furniture which make up the scientist's mind that are not in the focus recede into a kind of background knowledge (which is nonetheless semantically and logically connected with those concepts and beliefs within the focus), then the unity of mind and meaning could be preserved. However, if the mental furniture which does not inhabit the focus of attention simply disappears from the context altogether, it would seem as though the context is operating independently of the rest of the mind in a possibly undesirable way. Various issues arise which it is not possible to address fully here: how much the meanings of individual concepts within the focus depend upon the meanings of other concepts which have dropped out of the context altogether, for example. Even if this sort of sharply delineated contextualization were possible, one

wonders whether some particular belief and its associated concepts which are only a few steps removed from the present mental context of the scientist could be properly excluded from either the focus or the background of that context.

The objection, then, hangs in part on untangling these issues. Let us suppose, however, that these problems have been solved and that it is possible to inhabit a mental context of a sort amenable to the construction of the objection. Before proceeding, it is well to note what the objection seems to commit itself to in terms of the structure of (human) cognition. The objection requires a view of mental contexts which is much stronger than a mere rearrangement of focus and background, where the background includes *all* mental furniture not in the focus; it requires that the contents of such contexts are sharply delineated and can exclude large amounts of mental furniture even from the background. The objection also requires that *all* mentation that avoids the problem be context-bound. The context with the greatest possible scope would be that context which subsumes the whole mind and its mental furniture, a context identical with that favoured by mental and meaning holists. Since mentation within this latter context falls prey to the problem as I have described it, the objection must also carry in its train a commitment to the view that it cannot be the case that we ever acquire beliefs in this broadest possible context. In other words, the objection assumes a rejection of one of the central assumptions of many forms of holism.¹

The objection may have to advert to a variety of holism, however, because it may be the case that it is the consideration of issues within this broadest context which allows us to make decisions among conflicting claims adopted as true within particular, well-defined subcontexts. What happens when we broaden our attention from a particular

¹ Including, but not limited to, the holisms espoused by Quine and Davidson.

context (such as that of the scientist's investigation) to the holistic context? We must reintegrate the particulars into the whole. If we assume a simple case in which there are only two contexts, the particular context of the scientist's investigation, and 'the rest', reintegration could go very badly if the mental furniture excluded from the particular context includes a belief that conflicts, or more particularly a belief whose already deduced consequences conflict with some claim latterly taken to be true within the particular context. If the meanings of concepts are thoroughly interrelated in the holistic context, the attempt to reconcile the newly adopted beliefs acquired within the particular context with those (conflicting) beliefs already held could come to an abrupt halt, or at worst, cause the entire mental edifice to fall apart. If the process was merely halted, what will serve as the arbiter for deciding what to do with the beliefs acquired in the particular context? It might seem that the holistic context would have a mental priority which would require rejecting the conclusions reached in the particular context as either incorrect, corrupted, or meaningless, depending on how strong a view one adopts of mental and meaning holism.

A more complex case would involve the reintegration of two separate subcontexts, whose contents include conflicting beliefs, into the broader holistic context; but here, too, we find problems. How does reintegration of the two subcontexts into the holistic context proceed? We cannot assume that the reintegration proceeds serially. If that were the case, then the mental contents of whichever subcontext was reintegrated first would become incorporated into the mechanisms of judgment used to assess the admissibility of the believed contents of the second subcontext. *Ex hypothesi*, the contents of the second subcontext conflict with those of the first; and, because the first

now sits in judgment with respect to the second, the outcome is a foregone conclusion: the contents of the second subcontext will be rejected. Reintegration, then, must proceed simultaneously. But under what circumstances would this occur? Assuming that active cognition is limited to one context at a time, simultaneous reintegration would require that one ‘completed’ contextualized cognitive episode remain in stasis while cognitive activity ran its course in the second. While it may be possible that contextualized cognition could proceed in this way, it does not seem sensible to say that this would be a common or normal occurrence.

There may be a way to avoid these troublesome outcomes, but it will require us to describe a different kind of contextualization, or better, compartmentalization of cognition. I have in mind the sort of picture of cognition described by Christopher Cherniak in his book *Minimal Rationality*.

B. Compartmentalizing Cognition II

Cherniak’s proposal was developed as an attempt to get away from the idealizing tendency in the philosophy of mind, which fails to account for many of the ways in which humans in particular appear to reason. In the third chapter of his book, Cherniak discusses the inadequacy of the Quinean account of human belief systems. It is a consequence of the Quinean view that any reevaluation of one belief entails the reevaluation of others in the ‘web of belief’, in order to maintain consistency. Cherniak notes that Quine likely intended his account to be descriptively correct of human cognition; he then objects to this thesis in the following way:

But the belief systems of actual human beings do not inevitably and automatically readjust themselves appropriately in the way that Quine describes. The departures from Quine's idealization that we are concerned with here are certain types of forgetfulness; part of the human condition is in fact to fail to "make the connections" sometimes in a web of interconnected beliefs, to fall short of a synoptic view of one's belief system. For example, at least a decade before Fleming's discovery of penicillin, many microbiologists were aware that molds cause clear spots in bacterial cultures, and they knew that such a bare spot indicates no bacterial growth. Yet they did not consider that molds release an antibacterial agent....As a first "irrational reconstruction," we can say that the belief that molds cause bare spots seems to have been "filed" under the category of practical laboratory lore as information on undesirable contamination; the belief that a bare spot suggests inhibited bacterial growth seems to be in a different file, on microbiological theory. Thus the web of belief is not merely tangled; its fabric of sentences is "quilted" into a patchwork of relatively independent subsystems. Connections are less likely to be made between these subsets. The Quinean model does not take into account the basic organization of human memory.²

Cherniak centers his argument here around the idea of two related beliefs that do not conflict with each other. These two beliefs do not affect each other in the way that the Quinean model suggests that they should. If Quine's model were correct, we could conclude that Fleming's discovery should have been made much earlier than it was. Since it was not, there must be something wrong with the holistic model as it is employed by Quine. Cherniak concludes that human memory – and, one supposes by extension, human thought – is compartmentalized. Beliefs acquired in different contexts are not always cross-referenced and compared; nor are the consequences of their conjunction always computed. Cherniak's model holds some interest for us in the context of our present discussion on the possibility of acquiring conflicting beliefs.

Cherniak emphasizes the structural differences of two sorts of memory, in terminology familiar from traditional psychology. At any given moment, it is possible to

² Cherniak, C., *Minimal Rationality*, pp. 50-51.

distinguish “a ‘short-term’ or active memory and a ‘long-term’ memory...[t]he contents of the short-term memory correspond to what he is now thinking about, not necessarily consciously...; all other remembered information is in the long-term memory”.³ Further, Cherniak states that the

short term memory is conceived of as a working memory, not just a passive store. That is, unlike long-term memory, it has contents upon which operations can be performed, such as making deductive inferences from the activated beliefs there; in particular, the practical reasoning from beliefs and desires that results in action can only occur there.⁴

Long term memory appears to be analogous to a filing cabinet. Once material has been adjudged worthy of storage, it is relegated to a particular file in the cabinet, and remains there, (relatively) dormant. The contents of the file do not necessarily influence future judgments on material acquired through sensation or through cognition, except, perhaps, in those cases where the new material has already passed initial judgments concerning its admissibility to the corpus of knowledge, and is being considered for inclusion within a particular preexisting file. If human memory and cognition were structured in this way, it seems plausible that an individual could come to hold two conflicting beliefs without being aware that they conflict. However, the matter cannot be laid to rest at this point. Consistency is still a goal of rational cognitive systems, even if all that is required for making predictions of the behaviour of others is Cherniak’s minimal consistency requirement. Even if we assume that minimal consistency captures not only what is needed for prediction but actually describes the internal cognitive goals of a particular agent, Cherniak notes that it would be clearly irrational for an individual to say “Some

³ Ibid., p. 52.

⁴ Ibid.

sentence in $\{p\}$ is false, and p ".⁵ It would be irrational, on Cherniak's account, because in this case the conflicting beliefs (some sentence in $\{p\}$ is false, p) would both be considered to inhabit the short-term, activated, working memory described above. We can judge as rational an individual who holds conflicting beliefs only when those conflicting beliefs are both filed in long term memory, or when one of the conflicting beliefs is being considered in short term memory and the other is still filed in long term memory.

We need to elaborate Cherniak's model somewhat before we can determine whether it offers a way out of our problem concerning belief and belief revision. To begin with, Cherniak's model suggests that the short-term memory is strongly associated with the conscious mind.⁶ All active processing of beliefs takes place in the short-term memory: "only beliefs in short-term memory can be premises in reasoning; beliefs in long-term memory are inert – they do not interact with each other, and they do not affect behaviour".⁷ Among other things, this strongly suggests that beliefs stored in the long term memory do not participate in perception. This causes some tension for those inclined to adopt the view that perception is theory-laden, in the sense that beliefs previously acquired by an individual have some influence on that individual's interpretation and judgment of what is perceived. Is it possible to modify Cherniak's account, so as to allow the inert furniture of the long-term memory store to have some influence upon perception? *Prima facie*, there is nothing preventing us from taking such a free hand with the purposes of long-term memory; without too much effort, we could describe an abstract for a mechanism that would make sense of this. We could, for

⁵ Ibid., p. 51.

⁶ Although not identical to it, as noted in the quotation taken from p. 52, above.

⁷ Ibid., p. 59.

example, describe long-term memory as a kind of adjustable filter on incoming experience, the adjustments being made only by directives output from the short term memory. It is possible to consider a filter as a *passive* processor, and so thinking of the long-term memory as having this role to play in perception does not do too much violence to Cherniak's definition of the long term store as 'inert'.

However, if we remain true to the other constraints of Cherniak's account, this mechanism would not work. Given the possibility that two or more beliefs that make up the matrix of this filter could conflict with each other, the results of filtering inbound experience through it would be highly unpredictable and probably undesirable. Suppose that two such conflicting beliefs have been stored in long-term memory, and that the long-term memory is being used as a passive filter in the way just mentioned. In some instances, processing would proceed as normal, as the conflicting beliefs would not both be salient for the assimilation of such input. In other cases, where both conflicting previously stored beliefs are involved in the processing of new input, it is difficult to say how the processed results would be handled. For, after the primary processing performed by the long-term store, the results must be passed on (in this scenario) to the short-term store. Before we can assess how these results would be further processed by the short-term store, we would need to have at least some understanding of the mechanics of the filter that the long-term store is here considered to constitute.

C. Cherniak's Long Term Storage *qua* Filter

To begin with, as already mentioned, this use of the long-term store must be passive: there can be no conscious awareness of its actions. Let us examine three scenarios for the passive interpretative action of the long-term store. In each, the information is sequentially filtered through the various beliefs of the store. Many beliefs held there will have no role to play in the interpretation, but by hypothesis, two beliefs which conflict with each other do. In the first case, we imagine the sequential filtering resembles the action of polarizing lenses upon light; consequently, we should observe a canceling effect. The information would be filtered initially by one of the two conflicting beliefs, and all information not conforming to that prior belief would be eliminated from the information stream. The resulting, partially processed information would then pass through the filter of the second belief in the conflicting pair. This second belief would, in the same manner as the first, remove all information from the stream which did not conform to it. The resulting interpreted information stream would then no longer have any information left in it that could be associated with either of the two conflicting beliefs; in some cases, this might mean a total elimination of all information in the stream, in others, the resulting interpretation would suffer from a peculiar gap. These would both be odd results, and although it is difficult to grasp how the result might be apprehended by the short-term store upon receipt, it is safe to say that the resulting conscious experience would be quite abnormal.

The second case imagines the sequential action of each conflicting belief on the metaphor of pushing soft clay through two boards, arranged in series, with different

shapes cut into them. The first board extrudes a rectilinear oblong as the material passes through a square opening; the resulting object is then pushed through another board with a circular opening. The final result would resemble a squarish oblong with rounded edges on four of its axes; leaving the metaphor to one side, the result conforms neither to the first nor the second belief in the conflicting pair. The product would be of little use to the short term store receiving it. It cannot be understood by the short-term store as an instance of perception of the type associated with either belief: the short-term store has, in fact, no conceptual resources to draw on to make sense of this interpretation.

The third alternative is to imagine, again, a sequential processing of information; but in this case, each conflicting belief has the opportunity to work upon the stream of information in its original, unfiltered state. In this case, the information would be first filtered (by whatever method) in such a way as to conform to the constraints of the first conflicting belief; the information would then be separately *refiltered* in a manner that conforms to the second. The order in which the information gets processed by the two beliefs will determine the final result, which makes the interpretative process unacceptably arbitrary.

Cherniak's account, depending as it does upon the passive nature of the long-term store, makes it somewhat dubious for the long term memory store to serve the suggested filtering purpose in and of itself. There is the possibility that the *contents*, or some subset of the contents, of the long-term memory store could act in a filtering capacity, but we must reject the actual use of the long-term store itself as a filter as an unworkable elaboration of Cherniak's views. We must assume, then, that the short term store somehow performs the interpretation and judgment to which all inbound information is

subject before being stored. Despite the limited size resources⁸ of the short term store, it must have some mechanism for using the contents of the long term store (perhaps selectively) for the purposes of such interpretation and judgment. Whatever that mechanism is, it must be the case that the contents of the long term store (selectively) retrieved for assistance in the processing of incoming information temporarily commit the short term store to a particular state or configuration, under the auspices of which the processing of the new information actually takes place.

D. Cherniak's Short Term Storage *qua* Filter

As the short term store is subject to more stringent rationality requirements than the long term store, it seems reasonable to suggest that the short-term store *qua* information processor sits in judgment on the contents of the long term store, which has in any case no adjudication mechanism of its own. To that end, if the faculties of the short term store recognize a conflict among two items pulled from the long term store, it is the short term store which will determine which of the two conflicting beliefs, if any, will be retained, and which will be eliminated. Let us suppose that, in the course of interpreting some new inbound information, the short term store retrieves some particular file of beliefs which it judges relevant to the processing of the new information. Now suppose further that the store also retrieves another file of beliefs, for the same purpose, which happens to contain a belief which conflicts with the first. Insofar as the short term store occupies a particular state or configuration brought on by the application of the first long term memory file retrieved, what effects could we expect from the attempt to

⁸ Ibid., p. 52.

additionally apply the state or configuration required by the application of the conflicting beliefs stored in the second file retrieved? Since the contents of the two files conflict, the application of the second will yield an inconsistent state, which will produce garbage output. On the assumption that the short term store is so constructed as to accommodate the conflicting beliefs, and so could allow itself to occupy the resulting inconsistent state, there would not be any room, it would seem, for the short term store to then judge the garbage output *as garbage*. The store could presumably re-check whether the output is properly deduced from the inconsistent set formed by the union of the two belief sets, but of course it would find no problems *there*; an inconsistent set of premises will validly yield any output whatever. Given the catastrophic consequences this would have, this whole scenario seems unlikely. Instead, something else must take place when the second file of beliefs is retrieved from long term to short term memory. Let us suppose that the attempt to add the contents of the second file of beliefs from long term memory immediately produces an error condition. Presumably the first step following this event would be to isolate the conflict as much as possible. So let us suppose that the short term store initiates some diagnostic function which traces the cause of the conflict to two particular beliefs. If the short term store, having isolated the two problematic beliefs, regards them as beliefs, it will be in the position of having to adjudicate among what are by its own lights two *truths*. The extant assessment of these two items as being true might appear to prejudge the issue, as far as the short term store is concerned. Since both are true, the short term store must conclude that the apparent conflict must lie elsewhere.⁹ Of course, since the apparent conflict does *not* lie elsewhere, it is difficult to see how the

⁹ The problem is not mitigated by assuming that the short-term store has a built-in metabelief regarding the fallibility of the system whereby belief-fixation takes place. For the argument concerning this point, see the chapter on fallibilism, following.

short term store will be able to extricate itself. For in this case, the information being considered – two conflicting beliefs – is fully interpreted on the terms of the short term store; it is entirely unlike the assessment of incoming information which has yet to be interpreted, assessed, and stored; hence it does not appear possible for the short term store to simply reinterpret one of the conflicting beliefs in such a way as to dissolve the conflict. It would be an attempt to interpret an interpretation.

This suggests that the short term store cannot regard the two beliefs as beliefs, when it comes to adjudicate between them. One way in which it might avoid this is to perform its assessment at a higher order, under the auspices of a form of metalanguage – one that would allow the short term store to regard the two beliefs not as true, but as true-in-L, so to speak. In doing so, the short term store would free itself of the difficulties imposed by attempting to assess the conflicting beliefs as beliefs. What resources can this presumed higher-order reflective capacity of the short term store make use of in its assessment? Presumably it would have access to the long term store and may even be able to reconstruct the sequence of events which led the short term store to assess and file the offending beliefs as both true. Supposing that these higher-order reflections do not propose an obvious choice of which belief to keep and which to drop, the short term store will have to somehow eliminate both and force a reinvestigation. How does the short term store *qua* higher-order assessor handle this? It seems unlikely that it would just summarily wipe both beliefs from the long term store (and the short term store, where we must assume that the two beliefs are being held in stasis while this extra level of mentation is taking place). From the point of view of the phenomenal consciousness occupying or closely related to the short term store, which trundles along quite ignorant

of the debate raging at other levels, the sudden disappearance of two beliefs would be bound to produce some odd experiential discontinuities. It certainly does not seem a good fit with the normal experience of the recognition of error or conflict.

Perhaps the higher-order assessment functions of the short term store undertake an alteration of the conflicting beliefs, so that they remain in both the long term and short term store, but without the attached truth-attribution that characterizes beliefs. Again, from the point of view of phenomenal consciousness, this kind of adjustment would seem to engender experiential discontinuities inconsistent with (and possibly unintelligible to) our normal experience (“that’s odd...just a moment ago, p was true...and now it’s not”).¹⁰ The only way out of this peculiar position would be to make the higher-order functions part of those which come under the umbrella of processes which are apprehended by consciousness, rather than relegating them to a sort of unconscious background process of the short-term store, as has been tacitly supposed above.

Supposing this to be the case, we should ask whether these higher-level functions are active continuously, waiting to be made use of, or whether they only become active upon generation of an error condition. If they are continuously active, one would expect any beliefs passing through the processes of the short term store to be simultaneously apprehended as true (by way of the ‘basic’ processes) and also as just bare propositions with no truth attribution attached (by way of the ‘higher order’ processes). That scenario does not accord well with normal conscious experience. I am presently looking at a book

¹⁰ It can be objected that this should be rephrased as “just a moment ago, I held p true [i.e. believed p], and now I do not”. The rephrased version could be used to make a point similar to the original, but the original phrasing is not incorrect. It is used here to emphasize that, for an individual, believing p or holding p as true is equivalent to p ’s being true as far as that individual is concerned. In other words, in the circumstance being discussed, it would appear to the agent’s phenomenal consciousness that the proposition p had actually changed its truth-value. While such an unfortunate experience could conceivably be the result of some sort of psychotic episode and may well be a regular feature in the life of an Orwellian *goodthinker*, it does not, as mentioned above, accord well with our own normal experience.

on the philosophy of Daniel Dennett on my desk; as near as I can tell, my experience of this is just the simple case: there is a book about Dan Dennett on my desk. It does not seem to me that there is any second order apprehension to the effect that “‘there is a book on my desk about Dan Dennett’ is a syntactically valid proposition in L which does not conflict with any other occurrent apprehensions reducible to propositional form in L ” or similar. It seems more likely that the second order processes only ‘kick in’ once an error condition is generated by the collision within the short term store of two conflicting beliefs. There is a possibility that this latter scenario does accord with our own experience of the recognition of error or conflict among beliefs. Recognition of such inconsistencies does seem to result, for a short time, in an altered state of cognition in which we reflect upon certain beliefs as opposed to simply using them uncritically; when reflecting upon them, they are regarded only as bare propositions rather than beliefs. So let us suppose that the higher-order processes are activated only upon the occurrence of an error condition in the normal operations of the short term store, and let us suppose also that these newly activated processes are consciously apprehended as they do their work. In this way we avoid both the problem of experiential discontinuities and the problem of double-apprehension.

Now we may return to the question of how these higher processes deal with the conflict among beliefs. If the higher order processes find no obvious reason for rejecting either of the two conflicting beliefs, such as internal inconsistency in either belief, each must be reinvestigated. If this reinvestigation is conducted by the higher-order processes themselves, then their conclusions will be rendered in the language of those higher processes and not that of the normal level processes which usually govern the operations

of the short term store. These conclusions, rendered as they are in L^* rather than L , could not be stored directly in long term memory as is, or passed back to the normal level processes of the short term store; they would be unintelligible to those processes. Instead, the conclusions would need to be translated into object-level analogues. Formally speaking, this may or may not be possible. Even assuming that it is possible, it does not seem to accord with our experience of such situations. Resolution of errors of this sort appears from the perspective of the subject to be a seamless process, not requiring the translation of results at any point. The only point in the whole process of recognition *and* resolution of a problem in which a subject's experience might be discontinuous is the point at which recognition of conflict takes place (the initiation of the higher level processes invoked to deal with the error condition – a kind of semantic ascent), as noted above. However, after this point, the resolution process is continuous. This suggests that the actual resolution of such conflicts must be dealt with by the normal processes of the short term store. In that case, the role of the higher level processes would be only to strip away the truth attribution on each of the conflicting beliefs before returning them to the control of the basic processes, perhaps with an instruction to reevaluate both.

E. Cherniak-Style Cognition & Conflict

Cherniak's model was introduced in order to explore the possibility of escaping the central problem of this paper. Granting the argument which generates that problem, does Cherniak's model provide a way through? What we have argued here is that Cherniak's model allows for certain conditions under which the central problem does not

arise. In so doing, we have shown that the problem is not a *global* problem for Cherniak's model. However, it is of value to mention that there are conditions within that model which do allow the problem to arise. We turn now to the argument for this conclusion.

We have been considering, in the discussion above, how it would be possible for an individual to accept two conflicting beliefs. Recall that in the discussion of the preceding chapters, the rational revision of belief requires that two conflicting beliefs must both be present to the conscious mind of an individual, in order to generate the recognition of incompatibility which would force a decision between them. For reasons already given, the holistic view prevents adoption of two conflicting beliefs. Cherniak-type proposals make it possible to so adopt two conflicting beliefs, thereby making it further possible to achieve a cognitive condition which allows for choice among the two. Adoption of two conflicting beliefs is possible, however, only in those circumstances where the formation of the second belief in the pair is guided by a process of interpretation which does not use or in any way connect with the first belief, already stored in the long term store – in other words, only in those circumstances which make a direct appeal to the limited cognitive powers of the human mind. The information system which Cherniak's model describes is not itself holistic; however, holism does intrude upon Cherniak's model in a significant, albeit limited, way.

Within the confines of the short term store, we can consider the processing as proceeding in a manner analogous to the way holists imagine such information would be processed by the whole mind. This interpretation of the internal dynamic of the short term store is compatible with Cherniak's outline of its operations, and it seems reasonable to suggest that Cherniak himself conceives of the short term store in this way. Holists

generally agree that it would be irrational for an individual to believe two conflicting claims; to do so would violate the coherence requirement of holistic models. Cherniak appears to disagree with this view, but in fact he accepts it when it has been suitably qualified. It is irrational, on Cherniak's view, for an individual to hold two conflicting beliefs simultaneously within the confines of the short term store. The operations of the short term store, then, are bound by the same sort of coherence requirements which are characteristic of holistic models.

If this is a fair interpretation of Cherniak's views, we may consider the following situation. New information is input for processing to the short term store. The short term store duly proceeds with the processing, employing whatever interpretative mechanisms it has at its disposal, including but not limited to the selective retrieval of various files of information from the long term store. Once the information has been fully interpreted, it must be stored in a file of the long term store. The short term store decides upon a particular file as a suitable repository, and retrieves said file from the long term store in order to check the new information against the existing information previously stored. Suppose that said file contains information which conflicts with the new information to be integrated with it. Because the newly interpreted information has not yet acquired the status of a belief (as it has not yet been checked against the existing information retrieved from the long term store), it must be immediately rejected by the short term store as an inappropriate interpretation, because it fails the test of compatibility with the beliefs previously acquired and stored in the retrieved file. In other words, the argument made in previous chapters applies to certain circumstances which may arise in Cherniak's model, as much as it does to circumstances generally in the holistic model.

It is possible to raise an objection to the effect that the newly interpreted information is already a fully fledged belief before the short term store has decided in which file the information will be stored. In this case, the recognition of conflict with the beliefs in the retrieved file will simply return us to the scenario outlined above concerning the short term store's methods for grappling with conflicting beliefs, recognized as such, within the short term store. This objection only holds, however, if we drop coherence with previously accepted beliefs as a criterion for the transformation of a proposal, hypothesis, or candidate interpretation of information into a belief. If we wish to make use of such a criterion at all, we must conclude that the newly interpreted information is not a belief until it has been successfully checked against beliefs previously acquired and stored – if not all beliefs, then at the very least those beliefs which are stored in the particular file that the short term store has decided upon as its proper repository.

Is the loss of a coherency criterion in the process of belief fixation a serious loss? Cherniak clearly intends his proposal to avoid the *global* coherence requirement of Quinean models of the mind; however, we have argued that his model does not seek to do away with coherence altogether – the short term store, at the very least, is bound by coherence requirements with respect to its contents at any given time. Here we are considering an extension of the coherency requirements to include coherence between a belief-candidate and the contents of the preëxisting file which has been selected as the proper repository for that candidate. If Cherniak's model as applied here does not include the 'final stage' checking of a candidate for belief against at least some element of the existing long-term store, and in particular, a check against the contents of the particular

file of the long term store which is the intended storage location of the new candidate, then the model will have lost a particularly valuable form of error control.

Not so, it could be objected. The disposal of the ‘final stage’ check of course allows the interpreted belief-candidate to become a fully-fledged belief without any more processing; and now, when the short-term store deposits the belief into an existing file which already contains a conflicting belief, the short-term store will apprehend this conflict and compare the two problematic beliefs within the arena of the short-term store. But why should we assume that this will happen? Since the belief-candidate – now a belief – is no longer under examination, there is no motivation for the short term store to do anything apart from simply deposit the belief in a file. The prior existence of a conflicting belief in that file is of no material concern to the short-term store, and there is no reason that it should recognize any conflict. When I open my filing cabinet to put a copy of my latest tax return, recently mailed to the government, into the ‘Taxes’ folder, I do not open the file folder, take out all of its contents and examine them for potential informational conflicts with the new tax return; I just stuff the new papers in and close the cabinet.¹¹ There is no motivation, here, to imagine that the short-term store in its role as filing clerk is going to behave any differently. It is only later, when a retrieval of the whole file is required by some new circumstance – say, in the processing of more new information (or an audit) – that the short-term store may come to recognize its error. This would be a rather serious shortcoming, for now the system as a whole has the potential to admit many more conflicting pairs of beliefs than it might otherwise, as it seems to have no mechanism for preventing the incorporation of such conflicting beliefs at all. Whereas

¹¹ Although I may very well open and review all the contents of my ‘Taxes’ folder before I have signed and mailed the original return for this year to the government. That analogy would be appropriate if we were considering the filing of belief-candidates rather than fully-fledged beliefs.

Cherniak's programme was originally introduced here to show how it would be possible for such circumstances of storage (and later recognition) of conflicting beliefs to arise, it now seems that there is no way to prevent the storage of conflicting beliefs from arising in any particular case. Recognition of conflict would become a phenomenon *only* possible long after the fact of acceptance of a particular interpretation of experience. There would be no circumstance under which we would ever recognize a conflict immediately after the production and acceptance of a particular interpretation of experience.

We are now in a position to make a few concluding remarks on our excursion into Cherniak territory. The minimal rationality model is certainly an improvement, in many respects, on Quinean holism. On the assumption of the correctness of Cherniak's model it is possible to present solutions to the central problem of this paper for many of the scenarios in which it may arise. As we have seen, the model permits in some circumstances the acquisition of two (or more) conflicting beliefs, and we can construct mechanisms for dealing with such conflicts which respect the constraints of the model. Hence Cherniak's model shows us that rational revision is not always and everywhere impossible for a belief-based cognitive system. There is room also for the use of previously acquired beliefs in the interpretation of new input which does not require any assumption of mental holism, and which allows the long-term store to remain inert. Yet there are nonetheless circumstances in which the central problem arises, despite these improvements.

There is one further significant concern. The discussion has here endeavoured to show that Cherniak's model, suitably elaborated, can provide a solution for the central

problem in certain situations. In doing so, it demonstrates that the central problem of the paper is not a *global* problem for epistemologies which make use of the notion of belief as a cornerstone. However, the arguments used here to show the ways in which Cherniak's model can overcome the central problem impose somewhat unintuitive limitations on that model. As mentioned above, the success of Cherniak's model in overcoming the central problem is limited to those scenarios which directly depend on the limited information-processing capabilities of the human mind; yet the short-term store in Cherniak's system functions, within its boundaries, in ways which are relevantly similar to a Quinean holistic cognitive system. However, this feature of Cherniak's proposal is intentional. Cherniak's argument is not with the central tenet of Quinean holism, which is that an agent's extant beliefs play a role in the generation of new beliefs or the rational modification of existing beliefs for that agent. Rather, his argument is with the 'idealizing tendency' which Quinean models embody. That 'idealizing tendency' is captured by the assumption that *all* of an agent's beliefs will play a role in the generation or adjustment of beliefs in *every* case where generation or modification is warranted (for example, through the input of new information). It is important to note that Cherniak himself regards the Quinean model as an ideal. Were we constructed differently, with perhaps greater and more efficient cognitive powers, we could approach that ideal and even theoretically attain it. The problem which concerns us here is that in making use of the Cherniak model to (at least partially) solve – for belief-based cognitive systems – the central problem of this paper, we are bound to reject Quinean-style holism as an ideal. It is not that the Quinean model is difficult or even impossible to attain; it is an undesirable goal. Cherniak's model allows an individual agent to form and hold conflicting beliefs as

a consequence of the limited capacity of the short term store and the inefficiency, or haphazard nature, of the retrieval mechanisms employed to draw beliefs from the long term store. If the Quinean model is an ideal, then for Cherniak the ideal cognitive system would be one in which these limitations on the short-term store and its retrieval mechanisms would be eliminated: a situation in which the short term store was capacious enough to accommodate the entire contents of the long term store *at once*, and so make use of the entirety of the system's cognitive resources in the processing of new information. However, in harnessing Cherniak's model to our purposes (viz., the dissolution of the problem posed), such an ideal is no longer desirable. Were we to attain it, then we would lose the power to acquire two or more beliefs which conflict and so allow for the possibility of resolving such conflict by reconsidering each.

Our discussion of Cherniak's model provisionally granted the argument leading to the central problem which arises for Quinean holism, and partially overcomes that problem by appeal to the limited powers of human cognition. We are led to conclude that such holism does not represent an ideal for models of the mind. This is a rather counterintuitive result. If we accept that Cherniak's model represents a way of mitigating the severity of the central problem, we must reject Quinean holism not only as descriptively accurate but as an ideal as well. If we wish to continue to consider Quinean holism as an ideal, as Cherniak clearly does, we must withdraw the concession that the argument which leads to the central problem is cogent, or we must fully solve the problem that argument generates, and solve it in a way that preserves holism as an ideal. However, if we manage to solve the central problem for holism, then our excursus into Cherniak's model becomes superfluous for the purposes of this paper, except insofar as it

aids clarification of the issues involved – notwithstanding whatever other benefits might be derived from adopting Cherniak's model.

Before we return to a possible resolution of the central problem, it may be of value to explore its impact, or lack thereof, upon a general and extremely popular epistemological position – fallibilism. Some aspects of the central problem might be suspected of deriving from an assumed infallibilism; in demonstrating (or attempting to demonstrate) the ways in which the central problem applies to fallibilist positions, we will at once emphasize the general nature of the problem and provide further clarification of its structure.

Fallibilism

The scenarios recounted in the first chapters assume (or aspire to assume) nothing about the epistemological attitudes of the protagonist, our scientist. It would be well for us to examine whether our scientist could find herself in the predicament outlined in those chapters if she were a fallibilist. The problem as described arises most sharply for those we might call *pure* infallibilists, who regard any genuinely justified evaluation of a claim as *true* as being immune to error and hence also to revision. Pure infallibilists, however, are in notoriously short supply; more numerous are those who adhere to a kind of mixed-mode – or weak – infallibilism, which demands immunity to error with respect to some core claims to knowledge and yet allows fallibility with respect to others. If the central argument cannot address its own applicability to fallibilism, where the term covers both pure fallibilism and the fallibilistic aspects of a weak infallibilism, it possesses limited philosophical value.

The term ‘fallibilism’ itself should be elucidated before proceeding, as it becomes clear (upon closer examination) that not all fallibilist epistemologies are susceptible to the problem. This bodes well for the successful construction of an epistemological strategy which will resolve the problem, but may also reveal that those forms of fallibilism which avoid the central difficulty are not palatable to most individuals who would identify themselves as fallibilists.

A. Analysis I: Strict Fallibilism

The roots of fallibilism, it is generally agreed, are to be found in Peirce. As a term of art, fallibilism appears to have arisen to prominence relatively recently – perhaps in the last twenty or thirty years. There is no entry for the term, for example, in the Collier-Macmillan *Encyclopaedia of Philosophy* published in the late 1960s; it does, however, appear in A. R. Lacey's *A Dictionary of Philosophy*, published in 1976, where it garners the following one-line entry: "Doctrine that nothing or nothing about the world can be known for certain".¹ The term is more fully expressed in a passage written by Nicolas Rescher in the *Oxford Companion to Philosophy*, a recent cyclopaedia compiled by Ted Honderich and published in 1995:

Fallibilism. A philosophical doctrine regarding natural science – most closely associated with C.S. Peirce – which maintains that our scientific knowledge-claims are invariably vulnerable and may turn out to be false. On this view, scientific theories cannot be asserted as true categorically, but can only be maintained as having some probability of being true.²

This interpretation of the term may be unnecessarily narrow; although fallibilism has its origins in the philosophy of science, it is now widely used as a description of a general epistemological outlook. Nothing of significance, for our purposes, rests on this difference of scope, so it is possible to employ the relevant elements of Rescher's definition in the context of our discussion.

¹ Lacey, A. R., *A Dictionary of Philosophy*, Routledge & Kegan Paul, 1976.

² Rescher, Nicolas, entry in *The Oxford Companion to Philosophy*, Ted Honderich ed., Oxford University Press, 1995, p. 267.

For our purposes, it is the final two clauses of the passage quoted which are of most interest. The penultimate clause suggests that it may not be the case that a thoroughgoing fallibilist will ever assent to a claim (a theory) as true; that is, that a fallibilist will not take a proposition to be true, or in other words *believe* it. Note that the definition as stated also denies to fallibilists the position of believing a claim while also maintaining a metabelief to the effect that the assessment of any given proposition may be wrong. For Rescher, belief does not appear to enter the fallibilist position at all. Something like this interpretation of fallibilism may account for Karl Popper's adoption of E.M. Forster's dictum, "I do not believe in belief", although Sir Karl would never have agreed that we can speak in terms of some proposition as having some probability of being true, as the last clause in Rescher's description claims.

Nor, indeed, would Peirce. A discussion of Peirce's interpretation of probability would not be appropriate here, beyond mentioning that Peirce regarded probability in objective terms³ and not, in the manner of DeMorgan and Ramsey, as a method of assigning degrees of belief. Nor did Peirce condone the use of the probability calculus for the purpose of assigning 'some probability of being true' to a particular theory or proposition. In addition, although Peirce wrote extensively on the quantification of abductive inference in terms of probability, it is clear that he also countenanced abductive methods which did not rely on probability. Hence the fallibilism of Peirce does not qualify as a member of that species of fallibilism which is founded on probability altogether.

³ Peirce defends a version of the propensity interpretation of probability, and regarded the observation of frequencies as an objective method of access to these propensities, which are a real feature of nature. See Hookway, C., *Peirce*, Routledge & Kegan Paul, 1985, pp. 211-214.

Rescher's interpretation does not make it entirely clear as to whether fallibilism necessarily suggests a subjective probabilist epistemology or not; however, the point may be moot as his interpretation is not definitive. Fallibilism does not, in and of itself, entail a probabilistic outlook. It would be of help to examine some other attempts at a definition of fallibilism.

In Dancy and Sosa's *A Companion to Epistemology*, Anthony O'Hear also attempts an elucidation of the term:

Fallibilism. The idea that any of our opinions about the world or about anything else might turn out false. It is associated particularly with C.S. Peirce and Karl Popper, although many epistemologists would admit to being fallibilists to some degree because the underlying idea, that human beings are ever prone to error in their judgements, is clearly true. What separates fallibilists from other philosophers is the confidence each is ready to rest on what seem to be our epistemological successes. Even among fallibilists there are significant differences on this point. Both Peirce and Popper see human beings in biological terms, as organisms striving to adapt to their environment. But while Peirce sees the aims of knowledge-seeking as the removal of the invitation to doubt, an aim which will in the long run be successfully achieved by a convergence of all enquirers on the truth, Popper insists that we never have positive reasons to accept any belief. His fallibilism, indeed, is hard to distinguish from Scepticism.⁴

Note that O'Hear scrupulously avoids discussion of *probability* in his brief excursus into fallibilism. His silence on this matter is likely intentional; although probabilistic epistemologies are indeed a species of fallibilist epistemology, it does not do to make a general characterization of fallibilism in these terms, as O'Hear's discussion illustrates. It is precisely because of this that Rescher's interpretation of fallibilism is inadequate as a general account, whether he intended either subjective or objective probabilism as the proper interpretation of his definition.

⁴ O'Hear, Anthony, entry in *A Companion to Epistemology*, Jonathan Dancy and Ernest Sosa, eds., Blackwell, 1992, p. 138.

B. The Strict Fallibilist Escapes

A Popperian fallibilist, on O'Hear's analysis, would not be the sort of fallibilist for whom the difficulties we have been examining could arise. As O'Hear correctly notes, Popperians never positively endorse any claims⁵, whether those claims are made within the context of the natural sciences or the context of common sense reasoning. *A fortiori*, Popperians will not ever admit to taking a claim as true, or in other words admit to belief in any particular claim. The seriousness of the problem that is the focus of this essay would be largely dissipated if most fallibilists or fallibilist outlooks could be traced to Popper's understanding of the term; however, alignment with Popper's epistemological outlook is considered radical by most philosophers, and many explicitly reject the cogency of his position. It seems to follow that the majority of those who would count themselves as fallibilists employ an understanding of the term distinct from that of Popper.

That species of fallibilism, however, whose roots can be traced to Peirce, may not be so far removed from Popper's outlook as one might expect. O'Hear characterizes Peirce's outlook for the aims of enquiry in terms of the 'removal of the invitation to doubt', which does not in itself indicate any particular position on the role of belief.

Achieving settled opinion or belief, from which all the irritation of doubt has been removed, is "the sole end of enquiry" for Peirce in his earlier work. Having achieved settled opinion or belief, however, the motivation for further enquiry is removed. The early Peirce would never accept that one could believe some claim and yet qualify one's

⁵ With the *possible* exception of their own Popperian arguments and attacks on the positions of other philosophers.

belief by declaring its potential for being actually false. Believing a proposition *p* ought to remove, on Peirce's account, any motivation to append a qualification of potential error. However, this suggests that the early Peirce is actually something of an infallibilist with respect to belief.

Peirce later elaborated his views on fallibilism in the course of refining his discussion of the proper scientific method of enquiry. Peirce's 'scientific man' never believes his latest results or established theories:⁶

Nothing is *vital* for science; nothing can be. Its accepted propositions, therefore, are but opinions at most; and the whole list is provisional. The scientific man is not in the least wedded to his conclusions. He risks nothing upon them. He stands ready to abandon one or all as soon as experience opposes them.⁷

Belief has no place in science, which Peirce considers the domain of pure enquiry; it has only practical application. A practical man, such as an engineer, will make use of a scientific result by converting it into a belief.⁸ Presumably this means, for Peirce, that belief is required for action; the scientific man will not be acting on the intermediate results of his continuing inquiry. This does distinguish Peirce from Popper, whose position entails that belief is *not* a prerequisite for action; beliefs are not a necessary feature of a Popperian agent. However, Peirce's practical man is clearly no fallibilist. The scientific man who represents the fallibilist position in Peirce's later work has no use for belief; hence both for Peirce and for Popper, belief has no role to play in the fallibilist outlook.

⁶ Hookway, Christopher, *Peirce*, Routledge & Kegan Paul, 1985, p. 69.

⁷ *Collected Papers of Charles Sanders Peirce*, C. Hartshorne and P. Weiss eds., 1931-1935, volume 1, paragraph 635.

⁸ Hookway, Christopher, *Peirce*, Routledge & Kegan Paul, 1985, p. 69.

It turns out, then, that Peircean and Popperian fallibilism do not differ to the degree O'Hear's discussion suggests. Both see the ultimate aim of enquiry as the articulation of truth; for Popper, we proceed towards this goal – although we never have any means of determining whether we have reached it – by employing the method of falsificationism, eliminating to as great a degree as possible any false claims. For Peirce, we proceed by eliminating the invitation to doubt which (in his opinion) attaches naturally to any claim, and he specifies a number of means for (again, in his opinion) genuinely eliminating such doubt. For individual Peircean enquirers concerned with individual claims, doubt is always present and is in any case no bar to further enquiry. Hence fallibilists who trace their roots to Peirce will be as little susceptible as Popperian fallibilists are to the main problem of this essay.

If this is correct, we are left with some difficulties. To begin with, there are clearly differing opinions as to how fallibilism should be generally characterized; and we have only scratched the surface here. More importantly, the varieties of fallibilism cursorily examined above seem to indicate that fallibilists in general will not be susceptible to the problem outlined in the earlier chapters; this derives directly from the fact that the types of fallibilism discussed above seem to be united in their opinion of the inutility of belief in the domain of enquiry.

C. Analysis II: Common Fallibilism

Something is amiss. As O'Hear points out, many epistemologists would admit to being fallibilists; however, most epistemologists regularly discuss the problems of

justification, knowledge acquisition and change of opinion in terms of belief. There must be a contemporary understanding of fallibilism which is not covered by the definitions recounted above; we must dig deeper. Consider this definition from the online dictionary, FOLDOC⁹:

Fallibilism. Belief that some or all claims to knowledge could be mistaken. Although Peirce limited the application of fallibilism to the empirical statements of natural science, Quine extended it by challenging the notion that any proposition can be genuinely analytic. Unlike a skeptic, the fallibilist may not demand suspension of belief in the absence of certainty.

The last sentence is curious. It suggests that fallibilists do find a place for *belief* in their discussions of the methods and aims of reasoning. However, as we have just seen, the two progenitors of the fallibilist outlook would deny this to be the case. Perhaps the FOLDOC entry is an aberration.

The outlook proposed by the author of the FOLDOC entry is not unique, however. Consider these two passages from Dancy's *Introduction to Contemporary Epistemology*:

We know at any time that our belief-set is merely provisional. Revisions will be called for, and the need to revise may occur anywhere. This is a form of fallibilism.¹⁰

Another possible version of foundationalism holds that there are some beliefs...which are fully justified unless something arises to defeat their justification.¹¹

⁹ Although FOLDOC is an acronym for the phrase-title *Free On-line Dictionary of Computing*, it is a surprisingly rich (and apparently accurate) resource for technical philosophical terms.

¹⁰ Dancy, Jonathan, *An Introduction to Contemporary Epistemology*, Blackwell 1985, p. 118.

¹¹ *Ibid.*, p. 64.

Dancy is not endorsing the positions which either of these passages suggest. His book is intended as an objective introduction to epistemology, and aims only to clarify such terms as fallibilism, belief, and justification, and to explore the consequences of adopting particular epistemological positions. The first passage quoted above suggests that it is precisely *beliefs* that the fallibilist is concerned with; the second suggests that achieving belief is not the same as reaching Peirce's 'settled opinion', as beliefs can be seen as defeasible or in other words revisable.

D. The Role of Belief in Common Fallibilism

This brings into play a consideration which we will explore briefly: the semantic stability, for lack of a better phrase, of the term *belief* within the confines of common fallibilism. In the introduction to this essay, we touched briefly on the question of how to interpret the force of the attitude that the term *belief* is supposed to embody. Peirce explicitly regards belief as taking-to-be-true; many other philosophers adopt this understanding as well, but many appear to have no stated opinion. Dancy, in the work from which the above quotations are drawn, is silent on the structure of this particular attitude.

We need to decide whether the variety of fallibilism just described in section C is, in fact, a coherent fallibilism, without adverting to the authority of Dancy and other professional philosophers who appear to assume its coherence. If it should turn out that it is coherent, we can turn our attention to the question of whether our scientist could escape the predicament if her thought processes are guided by such a fallibilism.

Let us perform a little analysis on the position in question, and try to pump a few intuitions as well. Our scientist regards some proposition p as true (she believes p). She also believes that any of her beliefs may be incorrect, or in other words, false. Hence she can presumably believe that p may be false. Of course, for any contingent proposition p , p is possibly false; one does not need to be a fallibilist to accept that. Our scientist is making a stronger claim – that p may be actually false, even though she takes it to be true. There is a serious tension here. Having taken p to be true, how could our scientist allow that, after all, p may be false? From the subject's point of view, taking p to be true simply means that p is true. If one claims that a proposition p is true, why would one then want to claim at the same time that it might actually be false? Conversely, if one allows that some proposition p may be actually false, what could rationally motivate one to make the claim that it is in fact true?

What is curious is that if we rephrase the possible falsity of p in terms of error and judgment, it becomes fairly easy to extract the cogency of the position. Our scientist takes p to be true, but also accepts that she could be wrong about p , or more precisely, that her assessment – her judgment of p as true – could be erroneous. Her fallibilism, then, appeals to her own assessment of her powers of judgment. Recognizing that the methods she uses to make judgments are not infallible, she can deduce that for any judgment she makes concerning some proposition p , that her judgment in that particular case may be erroneous. One might think that this healthy skepticism regarding powers of judgment should spill over into a withholding of endorsement *in* judgment. In other words, the recognition of fallibility ought to issue in making only tentative claims – asserting only of any particular claim that it has not yet been refuted, rather than asserting its truth.

Certainly the ‘strict’ definitions of fallibilism discussed above seem to be in agreement on this point. However, we have seen above that some epistemologists have no difficulties in describing a fallibilism which allows beliefs, understood as a truth-regarding propositional attitude; the most direct statement of this view coming from the entry culled from the online dictionary FOLDOC. Favouring this view, there does not seem to be any necessary rational connexion between an acknowledgement of the limitations of one’s powers of judgment, and the withholding of endorsement of a claim in judgment. “It is my judgment”, our scientist may say, “that it is true that p . However, that judgment may be mistaken – it may be in fact false that p . Its falsity could be demonstrated by obtaining evidence of such-and-such.” There seems to be nothing intrinsically objectionable about such a statement. If there were, any interpretation of fallibilism would indeed be threatened with collapse into skepticism, or at the very least, into the hardline fallibilism of Popper.

Yet it is not altogether clear from this elucidation that the position is entirely stable. Intuitively, given the discussion of the paragraphs immediately preceding, we would want to be able to allow our scientist to maintain the substitutivity, *salve veritate*, of the declarative sentences “I judge it true that p ”, “I believe that p ”, “It is true that p ”, and “ p ”. However, in coupling any one of these options with the clause “but it may be actually false that p ”, the degree of discomfort intuitively felt is not the same. Insofar as there is a difference, the thesis of their intersubstitutivity cannot be maintained. But it is difficult to see which link can reasonably be broken. The most troublesome of the group is “It is true that p ”; to utter a sentence of the form “It is true that p , but it may be actually false that p ”, whether in a scientific or any other discursive context, would be *extremely*

odd. Upon hearing such an utterance, one might immediately want to ask the speaker why it is that s/he would make so bold as to claim the truth of p given their immediate qualification. It is not obvious that we may declare this sort of utterance to be straightforwardly contradictory; yet there is a genuine difficulty in imagining the sort of circumstance in which it may be reasonably said. The substitution of any of the other three sentences here presumed to be closely related to the bald claim “It is true that p ” mitigates, to a significant degree, the eccentricity of the utterance as a whole. It is not unreasonable to suppose, against what we agreed in the paragraph above, that the qualifying clause in each of these three other cases suggests or connotes a withholding of judgment, which connotation then bleeds back into the primary clause. If this is correct, it would suggest that there is no straightforward relationship of substitutivity between the bald claim “It is true that p ” and the other three; which further suggests in particular that believing some proposition p may not be regarded as taking that proposition to be true.

The intuition pumps we have yoked into this discussion are now groaning under the pressure we are exerting upon them, and their noises are not easily quelled. There is clearly a problem relating to the meaning of the term ‘belief’ within common fallibilism. However, our modest attempt to resolve it has the unfortunate consequence of undermining the plausibility of interpreting fallibilism as a position in which one can both believe p – where this is understood as a truth-taking attitude – and simultaneously acknowledge that p may be, in fact, false. This interpretation of fallibilism may not be stable, and under investigation prone to collapse into the interpretation derived from the strict definitions examined earlier in this chapter. Nonetheless, we have not managed to arrive at a decisive answer. At this point, let us reserve judgment on the matter of whether

the investigative path we have followed represents the only route through the problem of interpreting this variety of fallibilism; perhaps some detail has been missed which would allow an argument for the stability of a fallibilism which incorporates belief, where belief is understood as a truth-taking attitude.

E. Problems with Common Fallibilism

So we proceed with the assumption that the term *belief* has a steady semantic value within common fallibilism; that to believe some proposition is to take that proposition to be true. On that assumption, the fallibilism latterly characterized in the entry from FOLDOC and from Dancy's *Introduction* suggest that one can have a fallibilist attitude towards one's own beliefs. *Prima facie* – if only by way of an argument from authority - it appears to be a coherent fallibilist position to maintain that one could believe some claim and simultaneously acknowledge that the claim may embody an actual falsehood.¹²

Given these initial conditions, we can now ask the question: is such a fallibilism compatible with the problem situation outlined in the earlier chapters? In other words, could our scientist be such a fallibilist and still find herself in the predicament outlined there?

If we proceed as if the interpretation of fallibilism which incorporates belief is indeed cogent, we can allow that our scientist may not only believe some proposition *p*

¹² To reemphasize: as opposed to a possible falsehood. We are concerned here an individual claiming belief in a particular proposition acknowledging that that they could be wrong in their assessment with respect to that claim's relationship to the actual world, not acknowledging that the claim is *possibly* false (or in other words acknowledging that the claim *is* false in some *other* possible world).

and admit the possibility of p 's actual falsity, but that she may also specify the observational conditions under which p might be refuted. If in the course of an experiment our scientist has a perceptual experience which fits these observational conditions, she will be in a position to reject p . The question remains as to whether our scientist would ever be in a position to have perceptual experience of the sort required to achieve this crucial refutation.

Drawing on – and rehearsing – the argument outlined in earlier chapters, we can construct the following scenario. Our scientist is exposed to some raw, phenomenal data in the course of an investigation which, were they interpreted in perception in a particular way, would militate against p . *Ex hypothesi*, she regards p as true. As a fallibilist, she acknowledges that p may be actually false, and agrees that she would be rationally bound to revise her evaluation of p in the light of particular kinds of evidence, which she may have already specified to herself. As it stands, however, she regards p as true; she holds that belief fallibly, but not partially. That is, she does not, in this scenario, believe that p is merely likely or probably true. She may believe that she possesses a great deal of rational support for the truth of p , but that belief concerns her attitude towards her (subjective) justification for believing p . With respect to the proposition p itself, she regards that proposition as true.

Let us now turn to the potentially problematic interpretation of the raw data. If we imagine that her extant beliefs directly influence the interpretative process, as our earlier discussion attempted to show, we are faced with an immediate difficulty. An ideal interpretative perceptual processor would possess an utterly unbounded capacity for

generating possible interpretations of any given packet of information, but we are not so fortunate as to possess such an engine. In the case of the proposal to use an existing set of beliefs as the core of an interpretative system, the problem of limitation is particularly acute. The processing of all incoming information through the existing set of beliefs will produce only results compatible with that extant belief set. So the use of the existing set of beliefs in interpreting such information will yield no interpretations that conflict with any belief in that set. Given that we do frequently make interpretations of incoming information which conflict with theories or other ideas we already hold, it seems unlikely that the extant beliefs are adapted to this interpretative task. We require an alternative source if we are to proceed.

On the one hand, it seems that whatever set of beliefs our scientist already has would be the preferred candidate for processing information before delivering the final interpreted perceptions, because of the rich experiential texture that set of beliefs could conceivably provide. On the other hand, using exactly these resources seems too limiting; our scientist's experiences can never surprise her by yielding up for judgment anything which might conflict with her prior beliefs. There is at least one other candidate for interpreting inbound information: innate mechanisms, such as might underwrite what has been referred to earlier as bottom-up processing. However, while it seems very likely (even obvious) that innate mechanisms have some role to play in the process of interpretation in human cognition,¹³ the idea that such innate mechanisms are sophisticated enough to provide interpretations of incoming information rich enough to provide all the conceptual tags we would require of them is unlikely. As suggested

¹³ See, for example, David Marr's work on the analysis of vision systems.

earlier, we may well have an innate vision interpreter, for example, which takes care of distinguishing (in a particular incoming datum) foreground from background, edge from shadow, surface patterns, dimensionality, opacity and so on; and in so doing, conceptualizes, in a basic way, the raw phenomenal data streaming in from the eyes. But such an interpreter will not also provide the conceptual overlay of 'mug of hot tea'. That last part of the process of interpretation would need to be provided by some further interpretative mechanism whose procedures and conceptual resources are not fixed at birth. What is needed is an interpretative system which draws on and is responsive to beliefs and concepts acquired and adjusted in the course of normal experience.¹⁴ Innate mechanisms which provide the basic level of conceptualization in perception will not be adequate.

As before, we shall shift our discussion from the use of beliefs as participants in the process of interpreting inbound information, to the use of concepts. Whereas beliefs speak largely to how matters actually obtain from an individual's perspective, concepts speak to what is *possible*. For example, one might believe that the geometry of space is Riemannian, but one can conceive of how space could be (at least) either Euclidean or Lobachevskian. Let us suppose, then, that the interpretative mechanisms we are considering make use of concepts acquired in the course of normal experience – as well

¹⁴ As mentioned earlier, we can allow the possibility that such 'complex' concepts, acquired in the course of development of an individual cognitive system, may in fact be analysed into more primitive concepts which were given innately; nothing in the present discussion turns on whether or not this is the correct way of understanding the process of formation of concepts. Even one who endorses such an understanding of concept-formation, however, must grant that complex concepts are learned, and constitute a different sort of concepts from the primitive sort used natively in, for example, the primary bottom-up analysis of a visual perception into such components parts as foreground, background, size and distance. It is a distinctive feature of such learned, complex, concepts that different sets of them may be applied top-down to the same bottom-up processed information packet.

as, but subsequent to, whatever innate concepts we may have.¹⁵ This opens the possibility for the perceptual interpretative mechanisms of an individual to deliver not just a single interpretation of a particular stream of incoming information, but many. At another point in this essay, perceptual interpretative mechanisms were described as analogous to a filter on inbound experience. If any particular such analogy could be made with what is being proposed here, where concepts are used in the process of interpretation, the best (although by no means perfect) analogy might compare such interpretative mechanisms with the action of an optical prism. The number of alternative interpretations which such a mechanism might deliver will be limited only by the repertoire of concepts available to that individual.

As with the use of beliefs in interpretation, the variety of interpretations that can be generated by an individual's perceptual processor when that apparatus is guided by concepts is still restricted, but the restriction here is to what the individual regards as possible rather than her beliefs regarding how matters actually stand. With this in mind, it is quite possible that each instance of interpretative processing performed will yield a number of interpretations, some of which will be compatible with existing beliefs and some of which will not. As earlier noted, the resulting stream must be subjected to some sort of winnowing mechanism prior to acceptance, for it would not accord with our phenomenal experience to suggest that all of the possible interpretations will be perceived simultaneously. A question now arises concerning how it is that the perceptual processor winnows out the interpretation that will actually be judged acceptable.

¹⁵ The operation of the mechanisms actually responsible for the formation and acquisition of new concepts does not concern us here, although it may be said in passing that such mechanisms likely involve the operation of something we might call a faculty of imagination.

At this stage, an inspection mechanism is called into play. How does this portion of the system decide among the possible interpretations offered? It seems reasonable to suggest that the system will make at least part of its determination based on the consistency of each interpretation with what is already believed, although not necessarily *all* of what is currently believed; we remain sensitive to the arguments of Cherniak and others regarding the limitations of human cognitive capacities. Assuming, however, for the moment, that one of the beliefs in play for comparison is our scientist's prior belief that p , as well as any specific observational deductions she may have made with respect to p , then our scientist's inspection mechanism will find fault with any interpretation which conflicts with them. With respect to the remaining interpretations that do not obviously conflict with any beliefs in play (as there is no reason to assume that this first winnowing will leave only one candidate standing), other considerations may be brought to bear by the inspector to help decide amongst them – for example, considerations of simplicity, plausibility, and other desiderata which we need not specify here.

Leaving aside the rest of the process which eventually selects just one interpretation as acceptable, the argument suggests that if our scientist already believes p – that is, regards p as true – she is rationally barred from selecting an interpretation of the incoming information which militates against p ; in consequence, so long as she holds p true, she cannot also come to regard a potentially conflicting interpretation of the data as true; her interpretative processes will never deliver such an offending interpretation as a finished perception. Hence she is never in a position to hold the proposition p and the potentially conflicting interpretation on an equal footing (viz., as beliefs), under which evaluation the two could be seen to conflict and thereby motivate a rejection and

reevaluation of both. It is worth reemphasizing that this is the case *despite* the assumption that our scientist has already consciously derived the observational consequences of p , and believes that having a perceptual experience that conflicts with those observational consequences would constitute good reason for the rejection or reassessment of p . Such potentially conflicting interpretations can be generated by whatever conceptualizing mechanisms are used to prepare interpretations of inbound information for judgment, but will never actually come to be perceptual beliefs for our scientist, as her prior belief in p will require the inspection mechanism of her interpretative system to reject interpretations of incoming data which have the potential to conflict with that belief in p .

This leads to a possible objection: if the inspection mechanism is capable of recognizing potential conflict between a problematic interpretation of incoming information and the observational consequences of a believed proposition p , why would *that* particular recognition of potential conflict not motivate a reassessment of p ? The answer is that the problematic interpretation cannot be considered as actually conflicting with the belief in p or specific beliefs in the observational consequences of p ; the problematic interpretation – or any interpretation – is not itself a belief. As there is no conflict among beliefs, there is no motivation for reassessment of any existing beliefs. The problematic interpretations are not in any way challenging any existing beliefs. Let us examine the matter in more detail.

If the battery of tests performed by the inspection mechanism did *not* include a check for consistency with existing beliefs as part of a strategy for winnowing out a particular interpretation, we would not face the problem which we are here considering.

However, given that the interpretative mechanism here considered may generate multiple interpretations of a single experience simultaneously, it must be constrained by an inspection mechanism which allows only one interpretation to be delivered as the actual perceptual experience for the cognitive system of our scientist; and that inspection mechanism must have some rules of operation for making its selection. It would not do for our proposal to allow multiple interpretations in at once; that would not accord with normal perceptual experience.¹⁶ Gesturing to other potential methods that the inspector might use for the selection of one interpretation, such as considerations of simplicity or plausibility, seems unlikely to help. A test for plausibility would be a test for plausibility relative to an established body of beliefs; likewise for simplicity. Hence, at the very least a subset of an individual's extant body of beliefs will in one way or another be brought to bear in the evaluation of each interpretation in determining its admissibility to the corpus of beliefs.

In the case we are considering, a particular interpretation is submitted to the review of the inspector as a possibility and checked against a set of beliefs which include the belief in p ; these beliefs constitute a framework of how things stand as a matter of fact, for the individual concerned. This particular interpretation – this possibility – does not fit within that framework, despite the fact that the information processing system was able to generate the interpretation from the existing conceptual resources of that system. It is summarily rejected by the inspection mechanism and assessment proceeds to the

¹⁶ Even in those cases where more than one interpretation may pass all the tests of the perceptual processor (such as Joseph Jastrow's duck-rabbit 'rival schemata' image), human perception is incapable of *simultaneous* apprehension of two or more interpretations of a single information packet.

next interpretation made available for consideration. Since there is no *challenge* made to the beliefs brought to bear in the assessment, there is no problem of conflict and hence no motivation to revise any existing beliefs. The possession of a metabelief to the effect that all extant beliefs are potentially erroneous and subject to revision – what we might call the principle of fallibilism – does not have any role to play at this stage. It would become relevant only if any of the extant beliefs in the individual’s belief set were, in fact, being challenged by a perceptual belief which conflicted with one of them. Confronted with two beliefs which challenge each other, such a metabelief could have a crucial role to play. Because each belief is a claim that some particular proposition is true, a claim about how things are as a matter of fact, it may be by virtue of such a metabelief that a wedge can be opened to allow for reevaluation.¹⁷

For those who prefer to speak of error-trapping instructions (as opposed to metabeliefs), the same argument, suitably modified, will apply to such instructions as “In case of a perceived conflict between two extant beliefs, remove the truth-attribution from each and reevaluate as equal possibilities”. In the earlier discussion of the adoption of Cherniak’s proposal for a realistic modeling of cognition, we noted that proposals of this sort do allow for the possibility of adopting two or more conflicting beliefs. It is clear that in the scenarios allowed by such proposals, we can describe methods for the resolution of conflicts among beliefs; the use of error-trapping instructions, such as that described above, would be one way. As noted in that earlier discussion, these proposals do not offer a resolution for those situations in which the core problem yet arises – those situations being the ones in which newly interpreted information is evaluated through the

¹⁷ Not to suggest that other metabeliefs would not have roles to play here as well (or instead) – such as the metabelief that it cannot be the case that both A and $\sim A$ are true.

consideration of extant beliefs which would potentially conflict with those interpretations, were they to be adopted as beliefs. That is still the case here, even if we explicitly specify that we are considering a Cherniak-type proposal which incorporates the sort of fallibilism we have just been considering.

An important objection can be raised. It is agreed that *if* two beliefs were to be found in conflict by an individual mind, *then* on the basis of recognition of that conflict, a metabelief concerning the fallibility of all beliefs might come into play and/or an error-trapping procedure could be implemented; this procedure would employ methods for disposing of the conflict. We have here argued that in many cases, even when operating under a Cherniak-type model of cognition, such circumstances are rationally prevented from arising and so change of opinion in said cases becomes impossible. The argument is grounded in a particular view of the logic of interpretation in perception which seems in itself unobjectionable (apart from the bizarre result here argued for). That view claims that many possible interpretations of a particular potentially troublesome raw phenomenal apprehension are available to a given individual mind, and that each of those interpretations is assessed until one is found acceptable, acceptability here being constrained by what is already believed. Supposing that this view is granted, one might be inclined to wonder how *any* of those interpretations can manage to pass *all* of the tests posed by the extant set (or extant relevant subset) of that individual's beliefs – given that we would expect the particular packet of raw information in question, normally, to issue in a perception which would in fact conflict with one or another of those beliefs. Surely, one might object, there will be some extant beliefs which will make any of the

purportedly compatible alternative interpretations themselves unworthy of acceptance, since such alternatives will almost always involve some interpretative distortion.

What is wanted here is a general argument as to why it is that our scientist could generate some interpretation of her experience which would be judged acceptable relative to the limited subset of beliefs used for assessment of the various interpretations of that experience, in those circumstances where we wish to say that the normal or most natural interpretation ought to conflict with those beliefs.

It must be noted against the objection's demand for such a general argument that failure to satisfy it will not, in itself, demonstrate that cognition grounded in belief is after all a viable description of human cognition. At best, the success of the objection will show that there must be some problem in the way we have attempted to analyze the process of interpretation which yields perceptions (by showing that under our current logic of interpretation, no interpretation in the circumstances proposed above would be possible at all); at worst, the objection's success may be interpreted as a serious strike against the viability of belief-based models of cognition altogether (again, by showing that under our current logic of interpretation, no interpretation in the circumstances proposed above would be possible at all). Which path we might follow depends entirely upon whether there is an alternative model of the process of interpreting perceptions which radically departs from the family of proposals we have been examining, and whether that alternative is itself compatible with a belief-based model of cognition.

Nonetheless, attempting to answer the objection may well provide some further insight into the structure of the problem as a whole. It may be most agreeable to proceed

by cases, in order to elucidate the details of the objection, and afterwards examine the matter of whether or not a general argument can be offered against the objection.

Let us suppose that our scientist, alone in her laboratory, is testing the pH level for a particular liquid whose characteristics are unknown. The liquid is a new precipitate of a chemical reaction, and our scientist understands a body of theory which allows her to make a prediction as to the pH level of the precipitate; prior to testing the pH, she predicts that the liquid is alkaline. Let us also assume that she believes this prediction. The liquid has been collected in a container on the lab bench; the lighting in the room seems good, and in any case it is a sunny day and the blinds on the window are pulled back. A piece of litmus paper is selected from the sealed packet labeled 'litmus paper' and dipped into the container. As it happens, the pH of the precipitate is other than she believes: the precipitate is an acid. Notwithstanding her *belief* that the liquid is alkaline, we would normally expect our scientist to see the litmus paper turning a deep red, indicating the *in fact* highly acidic nature of the liquid; and we would expect her to infer from this perception that her belief in the alkalinity of the liquid is incorrect. However, her present belief that the liquid is alkaline interferes with the ability of her perceptual processor to produce as acceptable an experience as of *the litmus paper turning red*. So the processor must reject this interpretation and substitute an alternative.

However, the objection argues, we can imagine that the alternatives on offer will be similarly constrained by what she believes, and that these constraints will make all of the conceivable alternative interpretations inadmissible as well. Based on the circumstances of the experiment as articulated above, we can allow that our scientist

believes the following: that she is in good (normal) lighting conditions, that the paper is indeed litmus paper, that the liquid precipitate has not been swapped without her knowledge by a lab technician, that she is not dreaming, her senses are reliable, her mind is not under the influence of any psychotropic substances and she is thinking normally. Given these beliefs, what alternatives might be on offer that conflict with *none* of them?

Even within the constraints of the conceptual resources of our scientist, many interpretations of the visual experience are available to her perceptual processor that will not conflict with *any* of the proposed relevant subset of beliefs she uses to appraise each interpretation. But let us tighten the constraints just a little: in addition to the group of beliefs described in the paragraph above, we will want to specify that the processor *cannot* eliminate from *any* interpretation of the visual experience that it includes the feature of a reddening in her visual field which is coextensive with the saturated area of the shape in her visual field that is otherwise white. We introduce this restriction in order to respect the non-negotiability of certain features contributed by that part of the process of interpretation we have earlier described as ‘bottom-up’ processing. Note that this tightening still allows the possibility, *pace* (for the moment) the constraints imposed by the other relevant beliefs, that the perceptual processor can reject the interpretation of *the litmus paper turning red*. Supposing that just those beliefs, and no others, mentioned in the last paragraph constitute the relevant subset for the assessment of the acceptability of the various interpretations on offer for this visual experience, we can offer the following interpretation: the liquid precipitate saturating the litmus paper is *itself* turning red while the *paper* retains its virginal whiteness. This interpretation is one which could be generated purely by the same matrix of concepts used in processing the visual

information into forms susceptible of assessment by the inspection mechanism, including the first (and most obvious) interpretation offered above, that of *the litmus paper turning red*; we need not suppose that in order to generate the deviant interpretation that our scientist's perceptual processor need make any ad hoc addition to the concepts in that matrix. Now in such a case, it may well be that our scientist has no extant beliefs into which this deviant interpretation of the visual experience can readily enter a relationship; however, it seems fair to say that neither would this interpretation raise the possibility of a conflict with any of the relevant subset of beliefs being used by her inspector in determining its acceptability.

Supposing that our scientist finds this interpretation acceptable relative to the relevant subset of beliefs as specified, she would, of course, feel a strong consequent motivation to seek an explanation for the interpretation – a way to make the interpretation enter into relations with the rest of her beliefs (recall that having *accepted* the interpretation, it is now itself a perceptual belief). We can propose the following thesis as one our scientist may arrive at in order to achieve this integration: that the precipitate reacts to the substance, litmus, with which litmus paper is impregnated, and that the reaction causes the liquid itself to

- (a) redden upon exposure to a sufficient quantity of this substance,
- (b) immediately lose its alkalinity, settling at a neutral pH.

This thesis would explain the observed reddening without requiring a rejection of any of the beliefs in the relevant subset as described above. Point (b) is required in order to explain why there is not an observed purpling, which one might suppose to be the

outcome of the liquid precipitate both reacting to the litmus and thereby turning red (point (a)), and the litmus paper itself turning blue as it reacts to the alkalinity of the precipitate.

Certainly this would constitute an ad hoc maneuver, but offending the sensibilities of strict Popperians need not worry us here; as an individual who employs beliefs, our scientist is already far beyond the Popperian pale, and – *pace* Popper – ad hoc maneuvers are not always and everywhere undesirable (or irrational). What is of importance here is that our scientist has, by these methods, managed both to find an interpretation of her visual experience which completely accords with her extant beliefs, thereby allowing her to avoid the potential for conflict that might have been caused by her acceptance of the first interpretation offered in this scenario, that of *the litmus paper turning red*, and to integrate an alternative interpretation after accepting it into her extant belief set. Moreover, as argued above, our scientist can yet maintain that she is a fallibilist with respect to her beliefs, even if she had specified in advance of her experiment with the litmus paper that she would be willing to reject her belief in the alkalinity of the precipitate if she were confronted with an observation whose contents were that of *the litmus paper turning red*. For she has not been confronted with such an observation; her perceptual processor's inspection mechanism did not allow that particular interpretation to pass into the domain of her beliefs, and so the conditions of confrontation which might have brought her fallibilist metabeliefs into play were not achieved.

The conditions of the scenario can be further tightened without affecting the overall argument. Suppose that in addition to having a belief in the alkalinity of the

precipitate, our scientist has made the further step of deducing what her visual experience will be. In conjunction with other beliefs, she deduces that she will have a visual experience whose contents are at least partially constituted by *the litmus paper turning blue*. Having arrived at this belief, our scientist proceeds with her litmus paper test. Now there must be found an interpretation which accords with her belief that she will have a visual experience which includes *the litmus paper turning blue* even while observing a reddening in her visual field coextensive with the saturated area of the shape which is otherwise white. In this circumstance, the following interpretation is acceptable: the liquid precipitate saturating the litmus paper is itself turning red while the litmus paper turns blue. All that needs to be added are some ad hoc ‘bridging’ beliefs that will facilitate the integration of the accepted interpretation. We can here offer the following: the precipitate reacts to the substance, litmus, with which litmus paper is impregnated, and that the reaction causes the liquid itself to

- (a) redden upon exposure to a sufficient quantity of this substance;
- (b) become opaque, thereby concealing the reaction of the litmus itself to the precipitate;

and furthermore, that

- (c) the precipitate’s pH has not been altered by the reaction to the litmus, hence the litmus paper is itself turning blue.

We can continue to tighten the constraints on the scenario, but some interpretation will always be found acceptable by the inspection mechanism of the perceptual processor of

our scientist which will not conflict with any of her extant beliefs, and more particularly, any of her extant beliefs in the relevant subset used for assessment.

One might be inclined to wonder whether the argument made here holds even if we replace all of the specific beliefs we provisionally decided to include in the relevant subset with a non-specific blanketing belief to the effect that *all current conditions are normal*. Such a substitution would subsume all of the individual beliefs as described, concerning the lighting, the nature of the paper, her mental state, and the physical conditions of the experiment; nonetheless it does not seem that such a comprehensive belief would militate against the interpretation she finally adopted. Unless, that is, such a belief includes or implies some more particular belief which specifies that the most obvious interpretation on offer in any particular case must be accepted. Then we might argue about whether or not *the litmus paper turning red* constitutes the most obvious possible interpretation or not. However, it seems unlikely that the comprehensive belief that *all current conditions are normal* would itself imply that the most obvious interpretation must always be accepted. It is also unlikely that this specific belief would be a normatively *necessary* criterion that the inspection mechanism must use to decide among the potential interpretations of perceptual experience; things could go very badly if current conditions happened not to be normal.

One *might* argue that such a belief would have some *practical* advantage in the shoot-first, ask-questions-later conditions of the state of nature, and being so advantageous, has become hard-wired. This is highly debatable. That the mushrooms in *this* basket look a *lot* like the very tasty mushrooms we ate yesterday might yield an

obvious interpretation that *these* mushrooms *are* of the same type as those from yesterday. But let's be honest; we're much better off feeding a few mushrooms to the member of our tribe whom nobody likes and whom we allow to hang around expressly for this purpose anyway, and observing the results before we drop the rest of them into the evening stew. Until then, we will stick with the interpretation that the mushrooms in this basket do indeed look very similar to yesterday's, but will not venture any further. One might hope that our perceptual processors, being themselves in many senses evolved entities, would exercise the same level of discretion in their delivery of perceptual experience.

Let us take stock of what has been discussed so far. We have offered a few cases concerning some interpretations of a visual experience obtained during the course of a laboratory experiment. *Prima facie*, these cases offer evidence against the objection that *no* interpretation that can be offered in any given case will be acceptable to our scientist's post-interpretative inspection mechanism. Hence, if the objection is still to be considered problematic, we cannot regard it as a *general* objection which can be disproved by providing an exceptional case; the objection must be qualified. Our objector can respond to the cases offered so far in the following way: although the objection does not apply in these cases, they are implausibly underspecified. If our scientist's subset of beliefs included *all* of those beliefs which could be, or ought to be, deemed relevant to the assessment, no interpretation on offer could be found acceptable.

In being qualified in this way, the objection may run afoul of the restrictions imposed by the Cherniak-type models which we have so far argued may be able to avoid,

in some circumstances, the problem at hand. If the belief set required for assessment becomes too rich, it may exceed the plausible bounds of the short-term store. It would be unfortunate for this objection to succeed, only to find that it can be properly framed within a model which does not abide by Cherniak-type constraints. However, we shall waive these concerns for the time being, returning to them only if there turns out to be no general response to the (now qualified) objection being posed.

Let us return to the scene of our laboratory experiment, with the constraints set as they were in the second case proposed above. In that case, our scientist had arrived at a particular interpretation of her visual experience, and her inspector found this interpretation acceptable: the liquid precipitate saturating the litmus paper is itself turning red while the litmus paper turns blue. Our objector will want to block the acceptability of this interpretation, and may do so by pointing out the quite reasonable possibility that our scientist will include, as one of the relevant beliefs in her assessment set, the belief that the liquid precipitate itself will not change colour when exposed to litmus; that the liquid, whatever its other properties may be, is *areactive* with respect to litmus (or for that matter, the unsized paper within which the litmus is embedded). So another interpretation must be sought. As it happens, we could suggest that the visual experience be interpreted as one of the litmus paper turning blue while the wavelength of the reflected light is altered as it passes through the liquid precipitate. Bridging beliefs could be added to account for this interpretation by noting that the density of the liquid must be such that it alters the wavelength of light passing through it. Thus she would be able to maintain her beliefs in (a) the *areactivity* of the precipitate to litmus paper and (b) the

litmus paper turning blue, while still explaining the observed reddening embedded in her visual experience.

Our objector could now respond by saying that our scientist must have some prior belief in the density of the liquid precipitate which would block the interpretation that allows that the precipitate could alter the wavelength of light passing through it; but that belief would be insufficient on its own. Given that we may continue to introduce new possible explanations for the observed reddening, the objection requires that our scientist would need to have beliefs regarding *any* subjectively conceivable property of the liquid *and* of any other element of the experiment which could in any way alter the wavelength of the light passing through it on its way to her eyes. In order to block acceptance of the admittedly exotic interpretations which our scientist may conceive, the objection requires that our scientist possess specific beliefs on every conceivable property of every element involved, by her subjective measure, in the experiment and the environmental conditions of the experiment, up to and including properties of space and time; for presumably our objector will also want to block any interpretations which make use of localized spatial deformities which could alter the wavelength of light between the saturated paper and her eyes. Moreover, all of these specific beliefs must be included in the relevant subset of beliefs which will be used in the assessment of the interpretations offered. There are a number of problems which emerge if the objection is thus elaborated. As mentioned above, there may be a problem with satisfying the Cherniak constraints on what may be reasonably considered to constitute the contents of the short term store performing assessment on the interpretations offered. We continue to waive this particular concern. There is also the more general question as to whether fixity of belief concerning all of the

elements which could conceivably affect the interpretation of visual experience is itself plausible. Given the type of experiment being conducted, our scientist is presumably a chemist. It seems unreasonable to assume that she would have fixed beliefs concerning subatomic and cosmological matters which may conceivably effect the conditions and outcome of her experiment; we may grant to the objection that our original cases were implausibly underspecified, but the addition of all these additional beliefs makes the case in question implausibly overspecified.

Be that as it may, there is yet a further concern. On the assumption that our scientist is indeed gifted enough to possess specific beliefs on all of the subjectively conceivable properties of all of the elements which may potentially affect the experiment, it is doubly implausible that the beliefs she settles on concerning all of these elements imply that they will have no effect whatever on the experiment, which circumstance would be required in order to block any interpretation of her visual experience which attempted to exploit said properties in order to avoid potential conflict with existing beliefs. If we allow that our scientist is sensitive to the rich variety of elements which may be seen to play a role in the outcome of the experiment, it appears far more likely that additional beliefs concerning these elements would prevent her from making *any* deductions of the form we have considered in our earlier cases. Equipped with the galaxy of beliefs she is now presumed to hold regarding the circumstances of the experiment, it seems extremely unlikely that she would deduce, for example, that the litmus paper will turn blue. Even if she has particular beliefs that, *ceteris paribus*, allow her to deduce that the litmus paper will turn blue, because those beliefs are no longer the only ones involved in the deduction, she would at most be able to deduce only that “*ceteris paribus*, the

litmus paper will turn blue” rather than “the litmus paper will turn blue”, full stop.

Withdrawal of the *ceteris paribus* injunction would make it unlikely that she would reach any further beliefs regarding the litmus paper at all. This would put her halfway towards consistency with a fallibilism of the Peicean/Popperian type that rejects belief as a component of enquiry, and in so doing, avoids the problem posed by this essay.

We can strengthen the objection, and at the same time minimize the discomfort caused by the attribution of an implausibly large and overspecific belief-set, by tightening the relationship between the conceptual resources used in the generation of interpretations of experience and the relevant subset of beliefs used in the assessment of those interpretations. One could suggest that there might be some sort of correspondence relationship between the conceptual resources and the relevant subset of beliefs. Our objector could propose, for example, that the very concepts implemented in those beliefs used by our scientist in arriving at the deductions she made in earlier cases make up the conceptual matrix that will be used to interpret her later visual experience during the experiment. This would allow for the generation of a variety of interpretations, without requiring that the belief set must become implausibly large in order to provide a block against all subjectively conceivable interpretations, as the objection requires; it would also reintroduce the possibility of making the sorts of deductions which we were earlier considering in discussing the objection, such as the litmus paper turning blue.

Assuming these restrictions, is it still possible to generate an interpretation which would not be blocked by any belief in the relevant subset? Returning once again to the earlier case where the potentially problematic deduced belief has the content *the litmus*

paper turning blue, we may propose a different interpretation of our scientist's visual experience: there is a reddening in the visual field coextensive with the saturated area of an otherwise white shape. This interpretation will pass the assessment of the suitably restricted set of beliefs used for that purpose, for such an interpretation does not conflict with her belief that the litmus paper will turn blue; and yet, it makes no use of any concepts beyond those which could conceivably have been provided by that same set of beliefs for the construction of the conceptual matrix used in the generation of said interpretation. At this point, our scientist will want to compose some ad hoc bridging beliefs to make sense of the accepted interpretation, now a perceptual belief. At this point she is free to make use of any of the devices we previously employed, or, if any of those devices should be found to offer the potential for conflict with the beliefs in the relevant subset used for assessment, she may instead deploy some new bridging theory which makes use of other concepts not derivable from the beliefs in that subset.

What we have done is in effect pushed *all* of what might end up being troublesome in the various interpretations of her visual experience into the domain of the ad hoc theorizing which may take place after the fact (the *fact* being the acceptance of one of the interpretations); yet we have done this without modifying the perceptual processor in any way. As the influence (exerted by the subject's extant set of beliefs) upon the inspector built into the perceptual processing mechanisms continues to frustrate acceptance of an ever widening collection of potential interpretations of the experience in question, the inspector will only be able to pass as acceptable a shrinking set of ever more primitive interpretations, and in the limit, only one interpretation: that which is yielded directly from the bottom-up portion of the perceptual processing mechanism. In the case

we are studying, this would be the interpretation that there is an observed reddening in the visual field coextensive with an otherwise white shape. All conceptual enrichment of this primitive belief as delivered by the perceptual processor, will now take place after the fact of its acceptance; at this point, any conceptual resources may be drawn upon in the subject's efforts to integrate this perception with her extant beliefs.

In order to frustrate this particular tactic, our objector will need to specify that the set of concepts (and beliefs) employed in any ad hoc bridging theory which may be developed must also be constrained by the conceptual resources earlier made available for the generation of the interpretation. Such a specification, however, needs argument. On the face of it, there is no reason that ad hoc theorizing should respect such a constraint. One might even argue that the imposition of such a constraint does violence to the very notion of such theorizing. Popper detested ad hoc hypotheses, considering their employment irresponsible, precisely because such manoeuvres allowed an individual to hold onto all of his or her existing beliefs and yet still account for new experience which could potentially be interpreted as conflicting with those beliefs (thereby indicating a problem with one or more of them). Implied by this complaint is the idea that the ad hoc theorizer could employ any conceptual resources deemed expedient to serve the goal of relating new experience to the web of their existing beliefs. Popper's normative methodology aside, his understanding of what constitutes ad hoc theorizing seems eminently correct. Moreover, the tests required by our scientist's perceptual processor were directed expressly at the assessment of potential perceptual experiences within the context of the experiment. The fruits of ad hoc theorizing employed for the purpose of relating an acceptable interpretation of said new experience to extant beliefs need not be

constrained by the same tests. Such ad hoc theorizing is even licensed to imagine and subsequently employ new theories and new concepts, if the individual's existing conceptual resources are inadequate to the task. Certainly these ad hoc theories must conform to the restrictions imposed by the beliefs in the relevant subset in the sense that they may not conflict with any of those beliefs; yet the conceptual resources used in the generation of such theories may overreach those which ground the beliefs of that subset.

The objection can still be raised; the arguments made against the objection, however, still apply. Our scientist can construct ad hoc theories to reconcile the perceived reddening in her visual experience with her other beliefs; and our objector can still apply to block the candidacy of these ad hoc theories themselves for acceptance by specifying that our scientist should hold a particular belief which does the blocking. However, as the ad hoc theories can incorporate considerations not contained in the original subset of beliefs, our objector will have to allow both (a) that our scientist holds particular beliefs on all of the elements and their properties which could conceivably impact upon the outcome of the experiment, whether or not those elements and properties were included in the original subset, and (b) that our scientist holds beliefs on each of these which specify that these elements and properties are neutral with respect to the outcome of the experiment. Otherwise, our scientist could conceivably find some element or property which allowed her to construct an ad hoc theory to integrate the perceptual beliefs now produced by our highly restricted processor which would not conflict with any standing beliefs she possesses, and she would not, therefore, have to reevaluate any of those beliefs. However, as argued above, while (a) is itself unlikely, (b) is extremely implausible. It is far more likely that our scientist will have no fixed beliefs whatever on

the more exotic elements and properties which may, by her own lights, conceivably play a role in the outcome of the experiment. In which case, although we can no longer argue that she would be prevented from making the sorts of deductions we imputed to her in the earlier cases, as these deductions *would* be possible based on the limited subset of beliefs she gathered for the purposes of conducting the experiment, we *can* argue that where there is no fixity of belief, an ad hoc theory can be created. A significant proportion of the power of our objector's argument has been dissipated.

There is a possible exception to our argument in what we may refer to as a *limit case*; and it is in this case that the original objection as described may offer a genuine bone of contention. We will attempt a description and offer some possible outcomes here.

The limit case is intended to express the tightest possible constraints which may be plausibly applied to our earlier cases. In this case, the final turn of the screw would be the assumption that our scientist came to believe in advance of her experiment with the litmus paper that she would have a visual experience which included a *blueness* in her visual field coextensive with the area in that field occupied by the litmus paper. If we impute this belief to her and specify that it be included in the relevant subset of beliefs used in the assessment of the manifold possible interpretations of her visual experience, and furthermore couple this with the earlier injunction that any interpretation on offer *must* include a feature to be described as a reddening of an area of the visual field coextensive with an otherwise white shape, the objection acquires a remarkable force.

Assuming that this train of thought remains unobjectionable (to say nothing concerning its lucidity), we can conclude that our scientist may now be in a fairly serious

predicament; for none of the interpretations that can possibly be offered for assessment can be made to fit with the relevant subset of beliefs used in the assessment. However, while the troublesome belief is deducible from the relevant subset of her beliefs, it is extremely implausible that this particular deduction would be made. In our earlier cases, we suggested that our scientist might reasonably be expected to deduce from her extant relevant subset the further beliefs that (a) the liquid precipitate is alkaline, and (b) That the litmus paper will turn blue. The expectation is reasonable because these deductions bear on the objective outcome of the experiment. It seems peculiar to suggest that she would make a deduction concerning her own subjective phenomenal experience within the context of the experiment; although such a deduction is possible, it is not objectively related to the outcome of the experiment. If such a deduction were ever made by scientists in the course of laboratory work, it would be a rare occurrence.

Nonetheless, the objection certainly holds in this case; as a result, we cannot conclude that there is a general argument to be made against the objection. However, the objection holds only in this highly unusual limit case. Even in the absence of a general argument against the objection, we conclude that it does not provide an opening for a general method by which an agent who possesses beliefs may rationally come to change those beliefs in the light of new information, whether those beliefs are held fallibly or not. As mentioned before undertaking the task of answering the objection, it was remarked that its success would not, in and of itself, indicate the validity of a belief-based model for cognition. Rather, success would indicate either that the perceptual-processing models discussed so far were faulty, or that belief-based cognitive models face even more problems than have been otherwise supposed in this paper. The success of the objection

in the limit circumstance described above does not seem to indicate that there are any particular lessons to be drawn here; except that, in considering these issues, the counsels of Peirce and Popper may come to seem increasingly wise. However, insofar as exploring the objection has put pressure on the plausibility of the variety of perceptual processing mechanisms which we have so far discussed, it can serve as a motivation to seek an alternative. Such an alternative will be introduced and discussed in the chapter following, although, as will become clear, this alternative does not naturally lend itself to incorporation within a belief-based cognitive model.

The adoption of the common fallibilist outlook which allows that we can possess beliefs and admit that these beliefs are open to revision, then, does not appear to allow us to escape the central problem posed in this paper. It is worth reemphasizing that the argument here is not intended to show that we are as a matter of fact stuck with this problem, nor that fallibilism is a bankrupt position. As the earlier portion of this chapter attempted to demonstrate, there is little clear consensus on how the term ‘fallibilism’ ought to be construed; of the various interpretations which were considered as candidates, some certainly avoid the central problem by avoiding use of the term *belief*. The goal in the latter portion of the chapter was to determine whether a fallibilism which did make use of this term would, despite being a species of fallibilism, nonetheless be susceptible to the problem; this discussion suggests that it is susceptible.

It is not unfair to suggest that fallibilism was introduced as a philosophical position in order to placate a deep-rooted skepticism which many philosophers share. In particular, fallibilism shows promise as a strategy for solving the philosophical puzzle

posed by Hume's remarks on the vacillation of his attitude towards his beliefs. Whether or not fallibilism succeeds in addressing such concerns, the variety of fallibilism isolated in the foregoing discussion – so-called common fallibilism – does not appear to provide a solution for the problem now under investigation. In what follows, I shall develop an alternative strategy which may point the way towards solving the Humean problem by way of solving the problem about belief.

Evolution, Epistemology & Cognition

Having spent so much time attempting to make negative arguments about the role of belief in cognition, it seems only fair to give the reader some respite by describing something like a positive theory, and at the same time unveil more of the assumptions which have motivated me.

I. Evolution

In setting the scene for a presentation of this positive theory of cognition, it would be of use to begin with some discussion concerning the theory of evolution; for a significant portion of the characterization of cognition undertaken here will proceed by way of analogy with certain aspects of the mechanics of evolution.

It is clear that the Darwinian theory of evolution has captured the imagination of many thoughtful individuals whose interests have little direct relation to biology. In philosophy, for example, the theory of evolution has been pressed into service in attempts to elucidate or even support arguments for epistemological theses about the growth of knowledge and the structure of the mind.

The intent in the following discussion is to touch on the theory of evolution and place emphasis on some points which may be salient to our earlier discussions: first, the

teleological taint which may sometimes be discerned in the catch-phrases of evolutionary biology, and second, the persistence, as opposed to the proliferation, of genes across generations. Discussion will then turn to how the conclusions reached with respect to the theory of evolution as it applies to biology ramify to the evolution of knowledge (or its affiliates, such as thoughts and perceptions) in an individual cognitive system. That will lead to a discussion about some aspects of the structure of the model of cognition being offered to solve the main problem of this essay. Some of this will seem polemical and probably is; apologies are offered in advance.

I.A. Darwin, Teleology & Explanation

The idea of evolution by means of natural selection as Darwin expressed it may be a candidate for the greatest idea anyone ever had; as a scientific theory, however, it suffers from a number of shortcomings. For example, it has been noted that Darwin's theory of natural selection formulates no laws, and so yields, on its own, no empirical determinations.¹ This renders the Darwinian theory of evolution untestable; no particular empirical evidence can be described which has the potential to falsify it. Under Darwinian evolution, absolutely nothing need have turned out the way it has. One can use Darwinian evolution to *explain* why giraffes have long necks, but that explanation by itself will not explain why it is that giraffe evolution did not follow some other route. Instead of bearing progeny that had increasingly longer necks, giraffe precursors might have started producing litters of children with big claws, suitable for climbing trees to get

¹ Spencer attempted to formulate some laws in his own version of the theory, the so-called laws of 'differentiation' and 'integration'; these were vaguely formulated and hardly comparable to standard exemplars of physical law, such as the formula for Newtonian Universal Gravitation.

to the upper branches other animals couldn't reach. Or they might have produced children that had digestive mechanisms that could digest field grass, or insects, or even carrion. Or they might have produced children better able to withstand a broader range of temperatures than they themselves. The possibilities, while not infinite, are numerous; Darwin's theory does not predict that any particular outcome will obtain in any particular case. It is advisable to reserve judgment on the question of whether Darwinian evolution is properly characterized as a scientific theory or a metaphysical research programme; in any case, that issue will not be explored in this paper.

For our purposes, there is another, more relevant shortcoming. It arises perhaps partially because of Darwin's expression of his ideas, and partially because of the interpretations of those ideas as rendered by some of his supporters; this shortcoming is the infection of his theory by teleological notions. There is no question that Darwin intended his theory to have no such teleological taint; it is for this reason that his own theory of evolution was recognized as a genuine break with the explicitly teleological varieties of the theory of evolution championed by Lamarck, Paley, and even Hume. In his Herbert Spencer Lecture of 1961, Karl Popper remarked that one of the reasons Darwinism may be viewed as a significant advance on earlier explanations of evolution and speciation was the way in which it showed that "it is in principle possible to reduce teleology to causation" by explaining, in purely physical terms, these apparently teleological phenomena.² The spirit of Darwinian evolution is *not* best captured by the suggestion that its explanations are like Kipling's "just so" stories, which proposed that the animal protagonists had a particular problem to solve and developed some way of

² Popper, Karl, "Evolution and the Tree of Knowledge", from *Objective Knowledge An Evolutionary Approach*, Oxford University Press 1979., p. 267.

solving it.³ Animals, plants, and other phenotypic expressions of genes have no problems to solve insofar as they are phenotypic expression of genes; nor, for that matter, do genes themselves. Change – at both the genetic and environmental level – is entirely accidental; but from afar, variation, speciation and change may *look like* reactions to environmental pressures, and new phenotypic features may *appear* to possess adaptively advantageous functions.⁴ For many, this point will seem mundane, but it is worth emphasizing as there is a creeping tendency to overlook it. This tendency arises from what appears to be a widespread disposition – among biologists as well as philosophers – to use loaded terms when discussing evolutionary theory; terms such as problems, pressures, selection, and function, among others. This disposition is not altogether licensed by attempts that have been made to drain away the teleological blood which sustains the more common, uncritical usage of these terms.

Here's a story we could tell about the giraffe: a giraffe precursor was born at some time in the past with a genetic "defect" which phenotypically expressed itself as a longer neck. As it happens, this placed the food intake mechanism of the giraffe at a higher maximum elevation than any other ungulate competing for food resources in its

³ Or had their 'problem' solved for them; see Kipling, Rudyard, "How the Camel Got His Hump", *Just So Stories*, Doubleday © 1912.

⁴ Speaking, for a moment, in Neo-Darwinian terms, we recognize that gene expression is itself reactive to local environmental conditions. Recent studies of the *drosophila* genome have revealed, for example, extensive quantitative changes in the expression of various proteins encoded by the genome, depending on the developmental stage of the organism; presumably one could extrapolate from this research that the chromosome is executing a program which initiates new subroutines only when certain external conditions are met (external to the chromosome, i.e. within the cell walls; not the organism itself). Hence the *expression* of genes may be partially reactive to environmental conditions by proxy, if certain such conditions contribute to internal changes in the characteristics of the cells of the organism. The genes themselves, however, do not alter except under the impact of energetic radiation or through errors in copying during cell fission. See: Viktor Stolc, Zareen Gauhar, Christopher Mason, Gabor Halasz, Marinus F. van Batenburg, Scott A. Rifkin, Sujun Hua, Tine Herreman, Waraporn Tongprasit, Paolo Emilio Barbano, Harmen J. Bussemaker, Kevin P. White, "A Gene Expression Map for the Euchromatic Genome of *Drosophila melanogaster*", *Science Magazine*, Vol. 306, No. 5696.

local environment. This meant that a band of food resources was suddenly within reach of this one creature for which it did not have to compete (or at least, not to the same degree as its brothers and sisters), in addition to the food resources available to other members of its species; after all, the giraffe can bend its neck to reach leaves which are lower down. Access to more food resources, if food is otherwise scarce, promotes a better chance of good health, which means a better chance of surviving long enough to produce progeny, and possibly a better chance of surviving long enough to produce more progeny than might generally be expected given the local environment. Eventually, this variant might be expected, *ceteris paribus*, to do better than the members of its species who don't possess the defect which produced a long neck. That makes it reasonable to suppose (although not conclusively) that this particular gene type would eventually start to supplant the type which produces shorter necks within the giraffe-precursor gene pool. It may or may not have been the case that there was an inadequate food supply for those with shorter necks; although that condition provides one reason, it is not the only way to construct a story about why the long-necked varieties supplanted the short-necked varieties. Perhaps the leaves at the treetops produce an antibiotic which the leaves further down do not, and this protected the long-necked variety as a wasting disease worked its way through the population at some time in the past. A disease counts as an environmental pressure, if anything does; but the long neck does not count as a *reaction* to the presence of such a disease, any more than it would be a reaction to a scarcity of food resources. Another alternative: perhaps the precursors of giraffes have survived to this day, but can no longer interbreed with giraffes because of genetic drift. There are other stories we could construct here; Darwinian evolution does not itself favour any of

them above the others. Environments and environmental pressures certainly contribute to the explanatory stories we can tell, but the genes of organisms do not *react* to such pressures. The mechanisms which produce genetic variation are blind to the environment and environmental pressures with which phenotypic expressions of genes must cope.

Note that this explanatory story stays silent on the matter of problems to be solved, and eschews mention of the *function* of the phenotypic variation in question. What is the function of the giraffe's long neck? There is *none*; as it turns out, though, it conveniently places the possessing animal's mouth within reach of the upper branches of local food trees. One shouldn't assume that because some one thing facilitates or makes possible something else, that that thing therefore has that function. This heavy book on my desk, as it happens, does have a particular function: that of conveying, to those inclined and suitably equipped, its informational contents. Yesterday I used it to hammer a tack into a particularly stubborn bit of wall, but I'd never say that that was the book's function. Clearly this goes double for anything which has no function at all.

It seems a mistake to speak of function, design, solving problems, or indeed anything that even smacks of teleology when discussing evolution; at the very least, it should be done with *extreme* care. Even those inclined to agree with Darwinian principles can be found discussing the functions of various elements of organisms. For example, Jerry Fodor claims that "The function of the heart – the function that its design reflects – is to circulate the blood. There is no paradox in this, and nothing to affront even the most orthodox Darwinist scruples"⁵. Unfortunately for Fodor, there *is* room for an orthodox Darwinist to disagree; one might do so as follows. Ideally, one would wish to respond to Fodor by stating that, situated as it is within the bodies of most animal organisms, the

⁵ Fodor, Jerry, *In Critical Condition*, The MIT Press, 2000, p.191.

heart is a mechanism that *happens to* circulate the blood. Even if, in a charitable spirit, we allow the use of the word *function* in order to express the question of what some x does or what some x is for, the correct answer in the case of the heart would be: its function is to pump. What it is pumping, and what that pumped material does when it's not in the heart, is none of the heart's concern.

However, the second, more generous, response grants rather more than is prudent. It is wiser to simply avoid all such talk when dealing with evolution, as the danger of muddlement is too great. The term 'mechanism', coupled with suitable phrases or ancillary modifiers, should be able to take over a great majority of the uses to which the term *function* is applied in discussions of evolutionary biology.

Similar concerns can be brought to bear on the phrase *selecting for*. The very words themselves are suspect – 'selecting' and 'for'. Who is doing the selecting? And for what? If we can employ any terminology here, it may be best to don our Popperian hats and say that, for any given genetic trait of any contemporary organism, environmental pressures failed to eliminate it – or at least, failed to eliminate its phenotypic bearers (as a group). This locution seems to possess a great deal less of the taint of teleology than the phrase *selecting for*. However, it does raise the issue of how failure to eliminate a trait can explain how its phenotypic bearers not only survived but prospered, as in many given cases they have. Fortunately, a straightforward answer can be supplied: the propagation of a genetic trait through the phenotypic bearers of that trait is no more than the *repeated* failure of elimination across individual expressions of the genotype; possibly massive failure.

This is not intended as idle word-play. The tendency to describe evolutionary processes in terms of goal-directedness must be constantly monitored, as we appear to have an inborn tendency to ascribe functions and purposes to things we encounter. For practical purposes, this tendency is reasonably harmless. In the context of scientific and philosophical argument, however, it must be ruthlessly controlled. So much the worse for phrases like *select for*. If we do away with that phrase, we had better eliminate the use of the related phrase *survival of the fittest* as well. Unless, that is, we want to play a stipulation game, and stipulate the meaning of *fittest* as *whatever survives*; but that really would be idle word-play.

This brings us to the notorious phrase ‘natural selection’, from which the phrase *select for* is derived. This phrase has some hope of redemption. The presence of the word *selection* is something of an affront to the sensibilities that are being encouraged here, but coupled with the word *natural* its effect is mitigated. One studied definition of the term is “variation in reproductive success caused by genotypic variation”.⁶ Such definitions are to be preferred to more common dictionary definitions of the phrase, which often advert to terms such as ‘best adapted’, ‘fittest’, and related terms which carry a hint of goal-directedness.

In sum: evolution is not a goal directed activity, and genes neither solve problems nor react to environmental pressures. Genes react (and then only as individual genes, lodged in individual chromosomes) to cosmic rays and other physical processes which may directly affect DNA structures. Nature does not *select for* particular genes nor their phenotypic expressions, but only *fails to eliminate* phenotypic bearers of particular genes.

⁶ Lewontin, R. C., “The Units of Selection.” *Annual Review of Ecology and Systematics* 1:1-18, 1970.

It is entirely possible that the above efforts to expunge certain notions from discussions about Darwinism go too far, and in fact disembowel traditional orthodox Darwinism. On the supposition that this is the case, we shall refrain in further discussion from the use of the terms “Darwinian”, “Darwinist” and “Darwinism” except in those cases where the context might explicitly warrant it. Instead, we will make use only of the term “evolution” and related terms, where the meaning of such terms should be understood as constrained by the considerations offered above.

I.B. Persistence

Having framed matters in this way, we are in a good position to look at a portion of the theory of the origin of species which seems to be generally ignored. Most of the interest in evolutionary theory seems to revolve around the manner in which it manages to provide an explanatory framework for *change*; hence the emphasis in discussion on selection and environmental pressure. However, the explanation of speciation comprises not only the stories that can be constructed to explain change but also the stories which can be told about the lack of change. Darwin described the galvanizing influence of *An Essay on the Principle of Population* by Thomas Malthus on his own insights into the origin of species, which he came to see as characterized by a relentless struggle for survival. However, Darwin’s experiences in the Galapagos Islands showed him that genetic variation allows extant species of organism to exploit different ecological niches when those genetic differences were expressed as phenotypic differences. In these circumstances, genetic varieties do not compete for resources, and so could not be

described as struggling with one another. The environment exerts no pressures on the two varieties which would make them contestants in a zero-sum game; this can explain why both varieties may persist, rather than one supplanting the other, and in particular persist unchanged through millennia.

It is difficult to appreciate this consequence of the theory of evolution when one characterizes that theory in terms of the struggle to survive, or in terms of problem solving or reaction. Indeed, even too great an emphasis on the reasonably benign notions of natural selection and environmental pressure can interfere with the appreciation of the static element in evolution. The continued persistence of a particular genetic trait is generally best characterized in terms of failure of elimination. In order to understand the ramifications this characterization could have, it is useful to lay out the territory in a more exacting manner.

To begin with, let us define two types of genetic trait: essential and nonessential. An essential trait is one the removal of which will have a lethal impact (as opposed to a merely negative impact) on the survival of the organism which bears it. For example, the removal of the human genetic material which encodes for the phenotypic expression of a heart muscle will have a catastrophic result for the gestating foetus. The genetic material for the phenotypic expression of a heart muscle is, then, an essential trait for humans. Now it is of course possible that the genetic material that encodes for heart muscle could be replaced with different material which phenotypically expresses something which might substitute for a heart, such as multiple miniature hearts dotted around the circulatory system, or perhaps billions of tiny cilia in blood vessel walls which rhythmically fan the fluid around the circulatory system; or any one of a number of other

possible devices. In *this* sense the genetic material that encodes for heart tissue is not essential. *Ceteris paribus*, however, such material constitutes an essential trait. If we wanted to explain the persistence of such a trait across generations of an animal species, we would begin by saying that its sudden removal from the genetic code of a particular individual would result in almost certain death for that individual. Hence it would be unable to mature and spread its variant genetics into the population. If the genetic code for heart muscle were supplemented – within an individual organism - by code for a network of circulatory system pumps, which became larger and more powerful over a span of generations leading from that precursor organism, then at some point the code for traditional heart muscle would become nonessential; its sudden removal from an individual in one generation to the next might have some deleterious consequences for that individual, but not catastrophic ones. That individual may yet survive to produce its own progeny and the new code may then persist within the gene pool, and may conceivably supplant the old code for heart muscle under certain environmental circumstances. The word ‘essential’, then, as it is employed here, is a relative term.

A nonessential trait is one the removal of which does not have a lethal impact on the chances of survival of the bearer. Nonessential traits include those which have a negative, neutral or positive effect upon the chances of survival. Consider the case of a giraffe precursor whose genes encode for a longer neck than those of its brothers and sisters; this trait is nonessential to the individual organism’s survival. Removing the trait may make life more a struggle for the bearer organism, but would not ensure that the creature will perish. Its mother, after all, managed to live to maturity and bear children; hence the environment clearly allows for the success of the giraffe precursor without the

trait which generates a lengthened neck. This trait is a nonessential positive trait. An example of a nonessential neutral trait would be the code for the human appendix. In a small number of cases, this tissue can become a source of sepsis and kill the bearer organism. Generally speaking, however, its presence or absence makes no difference to the proliferation of the species. The category of nonessential negative traits should probably include deleterious modifications to existing traits, such as might produce a leg in the wrong location or bifurcating below the knee into two calves; on the assumption that such cases could be disputed, a clear example of a nonessential negative trait would be genetic code for an extra heel-bone, such that in the course of maturing it produces a bony process equal in weight and volume to twenty-five percent of the rest of the body of the organism to which it is attached. Just getting down to the cobbler's for a customized shoe would be a challenge for the individual possessing such a trait. Its removal would, far from being lethal, likely have a positive impact.

Having so specified the territory, let us return to the explanation of the persistence of traits. The persistence of essential traits is somewhat uninteresting and will not concern us here; such traits persist because their absence would be disastrous for an individual organism, and in all individual cases where an organism gestates without one of its essential traits, that organism is eliminated. What of the persistence of nonessential traits? In the case of a neutral trait, the explanation is straightforward. Neutral traits, such as the appendix, do not contribute to the well-being of the bearer organism, nor do they interfere with that well-being. Hence the organisms which bear such traits (and their children) are not, *ceteris paribus*, eliminated from the genetic pool of the species of which they are members.

Negative nonessential traits are not expected, generally speaking, to persist over many generations and so we shall not consider them as requiring explanation. Some may persist, if their influence on the chances of survival is not acute. The trait responsible for the architecture of the human spine comes to mind. The mechanics of human spinal columns are recognized as suboptimal relative to the sorts of employment they suffer at the hands of those who possess them; their weaknesses, however, are rather slight and do not often show themselves until the later stages of life, by which time the bearer organism has already had many opportunities to procreate and pass on its genes. That leaves only positive nonessential traits for consideration.

Recall the trait responsible for the phenotypic expression of a long neck in the giraffe precursor. This trait is positive, relative to the giraffe precursor's environment, because it gives the bearer organism access to a wider range of food resources than other members of its species. Does the fact that it is positive alone explain why this trait might persist from generation to generation? Its designation as positive might be used in an explanation of the *proliferation* of this variant type of organism within its species and perhaps, ultimately, the speciation of the variant. It does not explain why the trait persists. Rather, the trait persists because the bearer organism fails to be eliminated from the gene pool, just as in the case of the neutral trait. Any trait which does not harm the chances for survival of an individual organism will persist in the gene pool of the species to which that organism belongs. It may or may not become more widespread, that is, proliferate within the individual organisms which make up that species; that would depend on a number of factors. If we assume a random interaction and exchange of genetic material among bearer organisms within a species, then *ceteris paribus*, a positive trait should

spread through the gene pool over time. But the bare persistence of such traits from generation to generation depends entirely on the failure of those traits to harm the chances of survival for the individual bearer, and hence failure to enable the elimination of said bearers from the gene pool of the species. With respect to their persistence, such traits are neither *selected for* nor *selected against*.

II. Cognition

II.A. Evolution & Cognition

This being said, let us proceed to some discussion of the ways in which these considerations might be usefully applied to epistemological matters.

At the outset, we should note that as soon as we enter the domain of epistemology, the discourse allows for meaningful discussion of many of those elements which should be discouraged in the discourse of evolutionary biology: purpose, problem-solving, reaction. Humans undoubtedly have purposes, problems to solve, and react to new information. These factors do indeed weigh on the manner in which we develop and think about our ideas; however, the ways in which they influence our cognition can be seen to integrate with an evolutionary epistemological kernel.

The core notion, as in most evolutionary epistemological accounts, is to treat ideas in a way similar to that treatment we have given genes, in the section above. The term ‘idea’ is here to be used as a placeholder for mental content possessing propositional form – hence it comprises thoughts, whether entertained, accepted, or believed;

perceptions; judgments; and any other types of mental furniture which partake of concepts and can be rendered as propositions. We will not discuss the propagation of ideas from individual to individual, which would take us into a probably unprofitable discussion of the viability of Richard Dawkins' theory of *memes*. Rather, we will focus on the generation and persistence of ideas within the framework of an individual mind, and attempt to draw some conclusions about the epistemic valuations that individuals can make about such ideas. The analogy, then, will operate on two particular points: that *ideas* and affiliated notions such as *thoughts*, *perceptions*, and so on, are analogous to genes; and that a particular state-description of an individual cognitive system, or mind, is analogous to a generation in genetic biology. A state-description at time t_1 is analogous to one generation; a state-description at time t_2 is another. There may be sensed here an invitation to consider the relationships between ideas as causal, in the manner that the relationships among DNA molecules (or other potential gene-encoders) are causal. This is not our intention; we presume only logical relationships to obtain among ideas. This is clearly another point of disanalogy between genes and ideas, of which there are no doubt many more. Before going any further, then, it should be declared that for any comparable attributes for which there has not been a specific imputation of analogy, disanalogy should be assumed. This being said, let us proceed to the attempt to construct a workable cognitive model which respects the approved analogies here drawn between genes and ideas.

Our explanation will draw on some of the discussion lately made concerning evolution in addition to the analogy we wish to draw between genes and ideas. At this point, we have left the domain of belief; we are attempting to construct a theory of

cognition which does not make use of beliefs. Our motivation here is to overcome the largely intractable problem of accounting for rational change of opinion in a theory of cognition based on belief. In the construction of such a theory, we find a comparison with the behaviour of genes to be instructive, and in particular, the persistence (*not* the propagation or proliferation) of genes across generations.

Recall our earlier claim that genes persist across generations because their phenotypic expressions do not have a negative impact upon the survival of the bearer organism. We can tellingly rework this notion, for our current purposes, by making use of the word *challenge*: A gene may persist across generations because its phenotypic expression was not in any way *challenged* by the environment of the individual organism bearing that gene. For example, the gene whose phenotypic expression is the human appendix persists because the phenotypic expression of that gene was not attacked by a devastating infection in the majority of the individuals of the species bearing the gene. A gene may also persist if its phenotypic expression is challenged, but the challenge is overcome. For example, the appendix is so constructed that, in many cases, an infection does not ramify and create a condition of sepsis toxic to the organism as a whole; the appendix overcomes the challenge of most infections to which it is vulnerable. Of course, the appendix is aided in such contests by resources made available by other parts of the organism, such as the immune system.

If we apply this notion of *challenge* to the cognitive model we are developing here, we may find a way to characterize ideas that have been accepted and may constitute a basis for action or assertion. In the course of development of an individual mind, new ideas are generated from various sources, internal and external. Internally, new ideas are

generated by the imagination; external sources include, on the one hand, the absorption of propositional information encoded in natural languages, acquired from the speech and written words of others, and on the other hand, perceptual experience, which once fully processed itself possesses propositional form. These ideas are stored and accessed, compared, conjoined, consequences are computed, the consequences themselves stored, and so on. Some of these activities will be analogous to the solving of picture-puzzles: the activity of determining how to conjoin various pieces into a unified whole. In the course of such conjunctive activity, especially between hypotheses (absorbed from without or generated from within) and the propositions occasioned by perceptual experience, there will be challenges of adaptation. A particular hypothesis, coupled with some ancillary empirical hypotheses and initial conditions, may not readily fit with a particular piece of perceptual experience. As with a picture-puzzle, the cognitive system may possess a set of pre-defined 'pieces', that is, individual ideas, which may be used in the attempt to bridge the gap. Unlike a picture-puzzle game, however, the cognitive system is not limited in its resources to its box of precut pieces. It may design new ideas as its conceptual capacities allow, and it may choose to either add such new ideas to bridge a gap between hypothesis and perception, or to redesign or reinterpret either the hypothesis or the experience, respectively, in order to close that gap. In other words, both ad hoc manoeuvres and wholesale rejection and replacement of ideas are available as methods for overcoming the challenge of adapting ideas to each other. We discussed here the adaptation of a hypothesis to a perception, but such activities can also be applied among hypotheses as well.

In this discussion, no mention has been made of belief. The ideas (hypotheses, perceptions) which constitute the contents of this cognitive system simply persist within this system until they are challenged. Until that condition arises, these ideas may constitute a foundation for action and assertion. What about the ideas which may challenge the incumbents? New ideas may arise either through the workings of the imagination, or through the interpretation of experiences.⁷ In earlier chapters, we argued that ideas which were offered as interpretations of experience that would potentially clash with some extant beliefs within the system would be summarily rejected. This summary rejection was predicated on the epistemic evaluative inequality between the interpretation offered and the prior held belief. In the system we are now contemplating, this inequality has been eliminated. The ideas which occupy the roles of incumbents are not beliefs; they are nonetheless legitimate sources of action and cognition simply by virtue of their presence in the system. Ideas offered which may challenge these prior ideas have the same epistemic status, and so they do not offer a potential challenge; they offer an actual challenge, recognized as such. Once made, the challenge must be resolved by recourse to other cognitive resources we have not yet discussed. We shall not *recommend* any particular devices or methods for conflict resolution in what follows; our concerns are broader.

If the ideas persisting within a particular cognitive system are not beliefs, what epistemic status do they possess? We surely do not want to consider them as being merely *entertained* by the system in question; that which is merely entertained does not constitute a ground for action or assertion. Encapsulating any idea which is selected for

⁷ Such interpretative activity may also fall under the domain of the imagination, but nothing, for our present purposes, rests on the answer to this.

entertainment is a proviso to the effect that no judgment is being made upon it, as to its truth or falsity; by virtue of its status as entertained, such an idea does not have any epistemic value. In contrast, it is essential to a believed idea that it have an epistemic evaluation – namely, that it be taken as true, as expressing a fact about how things are.

The status of ideas persisting within the sort of cognitive system under discussion does not fit either of these categories. They are not encapsulated by any explicit proviso intended to express the non-applicability of judgment; yet neither is there any suggestion of commitment in their status. We shall name this status *acceptance*. An idea which has been accepted may or may not also have an epistemic evaluation attached to it; such evaluation is not essential to its status as accepted. In what follows, we shall attempt to clarify this admittedly peculiar status that ideas may possess.

II.B. Evaluation & Cognition

So far, we have introduced an analogy between genes and ideas with an eye to suggesting that the *use* of ideas in cognition is licensed⁸ by the mere uncontested presence, or persistence, of those ideas within a cognitive system. We further suggested that such ideas as persist within a cognitive system, as a corpus of ideas, possess no individual or collective epistemic priority over ideas which may be newly introduced into the system – by way of, for example, perception. The lack of evaluative inequality in the cognitive model slowly being outlined may strike the reader as somewhat alien. In further

⁸ This loaded term is here used with a conscious tip of the hat to those possessed of justificationist sympathies. In the most traditional of frameworks, an agent acquires the *right* to use a proposition as the basis for an argument or action by means of the provision, or potential for provision, of sufficient epistemic warrant. As the reader may have gathered, the view proposed here makes highly restricted use of such notions and so the word ‘license’ as used above retains only a vestigial form of its classical meaning.

elucidating the proposed model, let us begin by describing a hierarchy of cognitive systems which share certain fundamental features. The most basic of these systems will seem to bear very little resemblance to normal human cognition as it is commonly conceived; however, through the addition of certain sorts of refinement, the type of system proposed should become increasingly familiar.

Let us imagine, then, that we are considering a cognitive model which tries as much as possible to adhere to points of analogy with the theory of evolution in biology; it deals with conflict or challenge by employing a mechanism of random variation. We would consider, here, the existing corpus of ideas persisting within the system as being analogous to a conflation of the phenotypic expression of a genome and the genome itself, there being no distinction of this sort in the realm of thought. Certain modes of thinking of which this system is capable are readily recognizable to us; this system may, from its existing stock of ideas, deduce, add or eliminate further ideas.

Any cognitive system is partially constituted by the corpus of ideas which persists within its boundaries. However, the system itself is more than just that corpus; it comprises rules which pertain to the manipulation and interaction of those ideas, as well as structuring principles which dictate the internal organization of ideas and the mental faculties which play a role in their development. If this is a fair general description of a cognitive system, we may see that it allows us to interpret a cognitive system as constituting an *environment* which the corpus of ideas inhabits. The behaviour of the corpus is governed by the structure and rules of the system, considered as environment. One way in which the corpus can be affected by its environment is by the introduction of new ideas. Some new ideas may exert a kind of pressure on the corpus – a new idea may

be something which requires integration, or may motivate an evaluation of existing ideas within the corpus. Let us make an analogy between

- (1) the corpus of ideas within a cognitive system and the system itself, considered as environment;
- (2) a biological organism and its natural environment (nature).

We shall not make many references to this analogy in what follows, but offer it only as an aid to understanding. A certain subtlety attaches to this somewhat fanciful analogy; we must be sensitive to the plasticity of the roles which some of its elements may occupy. Just as ideas or thoughts persisting in a cognitive system may be subject to challenge and potential revision, so too may the rules and structuring principles of that system (although not necessarily *all*; this would depend upon the nature of the system). So we must allow that what we have described as the ‘corpus of ideas’ may include some of the rules and principles which organize that corpus; moreover, we must allow that what may be considered as open to the possibility of challenge (out of the elements of the system which are available to fill that role) can change. The introduction of this subtlety may perhaps unhelpfully blur the distinctions which this analogy intends to trade upon. However, the analogy is introduced only as a temporary clarificatory device; given the moderate repetitiousness of what follows, mentioning this caveat in advance rather than continually referring to it will help the analogy to fulfil its role.

In nature, there is no metalevel apparatus which examines the relationship between individual organisms and their environment, and which then attempts to tune or refine the composition of those organisms to eliminate the friction they may experience

in that environment.⁹ The resolution of conflict between organisms and some aspect of their local environment requires no such metalevel activity; conflicts are resolved, or not, by the interaction of physical processes alone. In our proposed cognitive system, however, matters are not quite the same. Conflicts between an existing corpus of ideas and a new, interloping idea will be recognized and resolved by metalevel cognitive functions. This is a feature which could reasonably be demanded of any cognitive system; differences appear when we consider the *methods* which may be used for conflict resolution.

The outline of the most basic possible procedure for conflict resolution is as follows. Suppose a new idea were to enter the system by way of some information channel, and that this new idea conflicts in some way with the preëxisting corpus of ideas. The system we are presently proposing – the first in our hierarchy of systems – will deal with the conflict by rejecting the entire corpus of ideas, and will then generate a new corpus from its internal resources of imagination. Such a system is undoubtedly crude, wasteful, and bears only an attenuated resemblance to any cognitive system we might wish to consider as descriptively accurate of human cognition. However, it is crude and wasteful in very much the same way that biological evolution is itself crude and wasteful. An organism which is defeated by a particular challenge simply dies; some other organism - possibly related but slightly different, or possibly altogether different from the first – will take its place, and may overcome the challenge if it is different in the necessary respects.

Crude as it may be, the system as proposed does embody a certain sort of cognitive virtue, in that it does not employ any form of non-demonstrative reasoning in

⁹ Unless we consider employing a divine agent, the use of which device we shall here eschew.

the adoption of ideas. There is no pretence that the adoption of any idea, or concatenation of ideas, is supported by empirical evidence or other sorts of premisses; indeed, no justificatory mechanisms (and the philosophical woes which attach to such mechanisms) are in operation at all. Ideas, or systems of ideas, are generated from internal resources, and face the tribunal of experience until a conflict arises; at that point, the whole is thrown out and a new set of ideas substituted, which may or may not itself encounter the same conflict.

Now let us refine the model somewhat, by allowing our system additional powers of scrutiny with respect to its own contents. In addition to recognizing conflict, let us suppose that the system envisaged may also determine various ideas in the corpus which may be the source of a recognized conflict; the elimination of any one of these sources would be sufficient for resolving the conflict in question. This somewhat more sophisticated system, rather than ejecting an entire corpus of ideas when conflict arises, may instead select some subset of ideas or just one idea for elimination. Let us suppose that this refined system selects at random from the various sources it may have determined as being potentially responsible, when a conflict is recognized. This would, in many ways, be more closely analogous to the random variation that is a key to the theory of evolution than the previous model; variation on the genetic level, after all, most frequently occurs as an adjustment to just one gene, rather than an entire chromosome – on the terms of our analogy, to a single idea rather than the entire corpus of ideas.

Such a cognitive system, like the one proposed earlier, would still be rather crude; but it would also continue to embody the sort of cognitive virtue of that earlier model, as it does not presume to use any non-demonstrative heuristic in order to decide which of

the potential sources of the conflict (this subset of ideas having been already isolated from the rest of the corpus) to eliminate; selection proceeds by random choice. It is not without justification that we might describe either of these two rather crude models of cognition as *Popperian* systems. In both cases, the general method of dealing with conflicts or falsifications of the system's theory of the world – the corpus of ideas it contains – is straight elimination and generation of alternatives.

And now we may add further refinements. Let us suppose that the system envisaged uses some criteria for making decisions among the various candidates for elimination, which criteria, while employed in demonstrative fashion, are not themselves demonstratively justified. As an example, we might suggest a criterion of simplicity, or the related criterion of minimal systemic disturbance. The choice of which potential source of the conflict to eliminate could be determined on the basis of such criteria, where the system could demonstrate to itself that some particular choice would require the least adjustment elsewhere, or could show that a particular choice is the simplest or most obvious one. At this point, our system should be slightly more recognizable as embodying something akin to the critical rationalism of Quine.

We may yet go further, and propose even more refined techniques and criteria for evaluating and deciding. In addition to mechanisms which affect the selection of ideas as targets of a recognized conflict, we may also suggest mechanisms for refining the sorts of ideas which are generated as candidate replacements. We may restrict, for example, the workings of the imaginative portion of the mind in various ways. In addition to mandating Quinean criteria such as simplicity, we may impose other criteria, parameters, or rules to influence and limit the generation of such ideas; we might even incorporate a

rule that forces the imagination to respect, as a constraint, the interloping idea which caused the cognitive friction to begin with.

To constrain the activity of the imagination in the way suggested above, or to posit selection criteria in the service of resolving cognitive friction, takes us in the direction of the domain of non-demonstrative techniques of reasoning. Such techniques need not be limited to such roles as we have expressed in the foregoing. We may employ techniques for evaluating interpretations of perceptual experience, for example, or for making ampliative inferences from a collection of interpreted perceptions. The more techniques we build into the mechanics of the type of system proposed, the closer such a model may come to resembling a more familiar system: that of human cognition. Human cognitive systems may regularly make enumerative inductions, evaluate ideas for their likelihood or probability, infer to the ‘best’ explanation, reject or approve ideas on the basis of such criteria as simplicity, and employ many more related and unrelated techniques for solving problems and attempting to improve understanding of their environment. All of these techniques are heuristics, either consciously deployed by the system or built into the system itself; such techniques happen to serve that system well when it is *urgent* that it come to some understanding of that environment rapidly, whether that understanding is accurate or not. These techniques, like the ideas that are manipulated by them, require no justification to license their use; just as ideas may be used in information processing simply because of their persistence in the system, so too these techniques may be used by virtue of their persistence alone. Other techniques may come to replace them, either through conscious replacement (in those cases where the techniques are based on rules consciously generated by the system), or by changes in the

structure of the system itself (through evolutionary processes which affect the biological foundations of the system). While such techniques make possible rapid and detailed adjustments to the corpus of ideas that the system contains, they are in no way essential to cognitive systems of the sort that we have been discussing in recent paragraphs. The only essential element in these systems is the ability to recognize conflict and the power to remove and replace sources of conflict. In the limit case, removal of a source of conflict would be the removal of the entire corpus of ideas, as the first, unrefined description of the model proposed. The additional techniques proposed for evaluating and choosing among ideas and interpretations of perceptions merely expedite, channel, and direct the process of alteration of the corpus. Moreover, their employment does not in any way guarantee that the alterations made with their assistance increase the overall verisimilitude of the system of ideas which comprise the corpus. The most that can be said of them is that they may assist the cognitive system in managing navigation through its local environment. Insofar as the cognitive system may have an internal goal of attaining truth, these techniques do not necessarily assist in the achievement of that goal – and may, in fact, hinder such an achievement.

Some of the techniques alluded to here play a role in making epistemic evaluations of ideas; others play a role in directing or restricting what sorts of ideas may be generated for addition to the corpus. The generation of new ideas addresses perceived absence; new ideas are introduced to replace other ideas rejected earlier, or to aid in the assimilation of new perceptions. The evaluation of ideas addresses problems of decision, where choice by random selection is not permitted. Evaluation is not required for decision, if random selection is permitted. Moreover, many familiar modes of cognition –

comparison, association, deduction, imagination and conflict resolution, to name a few – may be conducted entirely without recourse to epistemic evaluation of the ideas involved in those modes. Where circumstances are ideal, such that the cognitive system does not have to contend with such problems as, for example, mortality, no evaluative mechanisms are required.

However, such ideal circumstances do not well describe those of mortal cognitive systems. In any realistic scenario, practical exigencies will intrude upon the decision making process, making leisurely review and computation largely inconvenient, not to say dangerous or even impossible. Animal cognitive systems are quite good in serving the organisms to which they are attached as centers of processing of environmental information, and hence assist the organism in its attempts to survive. It is no wonder, then, that the types of cognitive system which have persisted and propagated across animal species have been those which refined, through evolution, various short-cut techniques of evaluating information and making decisions. They may be described as short-cut techniques because their methods go beyond deductive logic, and so their employment may be countenanced as a lower form of rationality than the employment of techniques which are founded purely in deduction. The study of such techniques may be classed as decision theory, which label is accurate so long as it includes the study of inductive logic and the systems proposed by such information theorists as Gärdenfors. We shall not intrude upon this area of discourse in our present discussion, but will offer some clarificatory remarks on the relationship between the contexts of recognition, evaluation, and decision.

In a belief-based system of the sort considered earlier in this work, one might consider the contexts of recognition and evaluation as fused. This becomes problematic, as we have argued earlier, if interpretations of perceptual experience or other ideas potentially conflict with prior beliefs; the potential for conflict occasions the summary rejection of said interpretations or ideas. When ideas are compared in an already evaluated state (e.g. as beliefs and possibilities), no actual challenge can be recognized and so there is no motivation to reevaluate.

In the non-belief-based cognitive system proposed in the foregoing, the situation is improved. In the case of two conflicting ideas, one persisting within the system and the other freshly generated through the mechanisms of experience and/or the imagination, the system begins by recognizing the conflict. The next step is to impose a primary evaluation on both. Here, the system might employ one or another of the short-cut techniques alluded to earlier. One technique might be the application of a form of probability-evaluator. The techniques would be applied to both ideas; the resulting evaluation would allow the system to move to, or may form part of, the context of decision, where further techniques may be employed, perhaps a logic of belief revision of the sort recommended by Gärdenfors and his disciples. The end result will be either an accommodation of both ideas, or rejection of one or the other.

It may be derived from the foregoing that we are suggesting the notion of ideas persisting within a cognitive system as being considered in a non-evaluated mode, for such purposes as general cognition and information processing. As mentioned earlier, this non-evaluative status is what we intend by the term *acceptance*. Possessing an evaluation is not, in this model, an essential or necessary characteristic of an idea which persists in

the system (i.e., is *accepted*) and which may form the basis for action or cognition; some ideas may have such an evaluation, and some may not. One category of ideas which would generally not have any such evaluation in the first stages of their existence would be those which constitute a fully interpreted perception. When everything is proceeding 'according to plan', so to speak, any evaluations which may be attached to any of the ideas involved in current processing have no role to play. Only when conflict is recognized are these evaluations examined and used for decision. Recognition of a conflict triggers an evaluation procedure. If each of the ideas were already associated with some sort of evaluation, the evaluation procedure could proceed from those prior evaluations. It may be the case that neither party to the conflict has such an associated evaluation prior to the current conflict, as such evaluations are, in the proposed system, only generated when an evaluation procedure is specifically triggered by a cognitive event such as a conflict or an independently motivated suspicion of doubt. The system would need to generate initial evaluations in order to proceed. It is also possible that only one of the ideas involved in the conflict possesses a prior evaluation; in this case, an evaluation would need to be made of the idea lacking one.

As mentioned above, in the normal course of processing information where conflicts do not arise, no evaluation of the ideas used in such processing is required. The mere persistence of ideas within an information processing system is sufficient to license their use by that system. The fact of their persistence, like the persistence of genes across generations of individual organisms in a species, is the outcome of a historical process. In the case of genes, the process is causal; in the case of ideas, the process is, by and large, logical. The ideas used in such 'normal' circumstances need not be believed nor have any

other positive epistemic evaluation attributed, nor even have a justification attached to them in order to license their use, although such justifications – which could contribute to an epistemic evaluation - may be in fact attached.

At least one further clarification of the notion of an accepted idea is required here. We have set out our explanation of the internal dynamics of this cognitive model in terms of a corpus of ideas on the one hand, and new individual ideas on the other. It is reasonable to characterize the corpus of persisting, accepted ideas as being at least partially systematized. Why partially as opposed to wholly, we will address in the section immediately following. New inputs or new ideas – for example, the fully interpreted perceptions – are initially outside this corpus. It may be a natural reaction to conclude that, being outside the corpus, these new ideas are not themselves accepted, and hence there is a significant inequality in the status of ideas within the corpus and those without – an inequality which could regenerate the problems discussed with reference to belief-based systems. However, this is not what we intend. Those new ideas which have entered the cognitive system but are yet outside the corpus of ideas which persists within it, are themselves accepted just as much as the ideas within that corpus. The equality of status allows for the recognition of actual conflict. Even if this is granted, our way of describing the internal dynamic of the system so far remains vulnerable to criticism along similar lines. By dividing the ideas within the system into those which are *inside* and those which are *outside* the proposed corpus, an argument can always be made that there is an evaluative inequality of some sort to be found which might regenerate the central problem, even if the inequality is not strictly speaking one in *epistemic* evaluation.

Our present characterization of the internal dynamic of the proposed cognitive system, however, is not to be taken literally. As mentioned earlier, we introduced the device of ideas being within or without the corpus in order to make our description of the dynamic more perspicuous. At this point, it would be wise to kick away the ladder. There is no real division of inside and outside within the proposed system; the explanatory devices used have no intended meaning apart from the separation of new ideas from previously extant ideas for purposes of exposition alone. Once an idea has been accepted by the system, it is already within the corpus; there is no outside.

II.C. Partial Systematization, Conflict, and the Will to Systematicity

A new idea brought into the corpus is accepted just as much as any other idea in that corpus. If the new idea does not conflict with any other member idea (or, as Cherniak would have it, any member of a suitable subset of ideas) – and here we mean *conflict* in the sense of logical inconsistency – then there are no issues to address, and no cognitive work to be done, except perhaps a modicum of integrative activity; the amount of such activity would depend entirely upon the level of systematicity of ideas that the system internally requires of itself. My perception of a teacup on the desk next to my keyboard presents not the slightest hint of cognitive friction relative to the other ideas I hold at present about my current circumstances, and so the idea that there is such a cup on my desk is simply absorbed without remark. The sudden and unexpected appearance of a storm of white flakes in the window beyond my desk, in early summer, might require the minor cognitive labour of theorizing that a gust of wind has blown the cherry blossom

petals off of the tree in my neighbor's yard – not because the sight of white petals causes any cognitive friction with my extant ideas, but because such a sight is not expected; it is not derivable, or at least not obviously derivable, from my extant ideas. I am in no way bound to make such efforts of integration; as it happens, my mind appears to possess a tendency to systematize. Despite this tendency, a little effort allows me to defer the integration of new information, if I am so inclined.

It seems possible that there could be individuals for whom the default tendency is *not* to integrate new information in this way, or integrate new information only in some limited domains of ideas and not others. Insofar as an individual does attempt such integration, their thinking is systematized, or has a systematizing tendency, in that domain. Human cognitive systems are almost undoubtedly only partially systematized in this way: some domains are, and some are not. We shall produce some examples of unsystematized cognition below.

It should be noted beforehand that the acceptance of certain methods or ideas which *prima facie* conflict with others within the corpus, and where such conflict does not occasion attempts to reject and revise or reassess some of those ideas, does not always fall under this notion of partial systematicity. There are scientific contexts, for example, where a discredited theory or method is nonetheless used to make predictions. Newtonian universal gravitation, notoriously, is still used by those involved in the science and technology of spaceflight. There are almost certainly some individuals involved in those endeavours who, having grasped both Einsteinian relativity and Newtonian mechanics, yet have no difficulty in employing the equations of Newton in their calculations of orbits, payload, escape velocity, and so on, despite knowing that

Einstein's ideas have supplanted those of Newton as generally approved descriptions of physical laws. This does not mean, however, that these individuals refuse to integrate the two theoretical systems, or fail to recognize their incompatibility. Rather, their use and understanding of Newtonian theories is enveloped in a set of caveats and 'notwithstanding' clauses, which have the net effect of licensing¹⁰ the use of theories in practical contexts, for purposes of expediency. Since Newtonian equations are 'good enough' for the uses to which they are put in such contexts, they may be used in lieu of the more complex equations of Einstein. In this way, the theories of Newton are fully integrated in the cognitive systems of those individuals who make use of them in these practical ways, in a manner which prevents the conflict that might arise were these ideas to coexist without qualification in a cognitive system which also accepted the theories of Einstein. Although the individual concerned may be regarded as *accepting* both Einsteinian relativity and Newtonian universal gravitation, their acceptance of the latter theory may be considered as being more restricted in scope than the former.

For a more plausible example of genuine lack of systematicity, we may turn to an anthropological study made by Evans-Pritchard, in his work *Witchcraft, Oracles and Magic Among the Azande*. Evans-Pritchard attempted, over the course of a number of years, to comprehend the system of ideas that Zande individuals possessed with respect to witchcraft. One conclusion he appears to reach in his work is that the Azande do not have a well-systematized body of ideas associated with witchcraft. In a chapter entitled "Are Witches Conscious Agents?", Evans-Pritchard attempts to determine, through questioning, whether or not the Azande think that witches are conscious of their powers and act with malicious intent. What he apparently discovers is that, while the Azande

¹⁰ The sense of *license* here employed is that referred to in an earlier footnote.

agree that all witches are always conscious of their status as witches, conscious of their specific abilities, act with intent, and collude with other witches (whom they are naturally able to identify without special signals), this attitude is not self-reflexively applied. If a poison-oracle identifies a particular man as a witch, that accused individual may come to accept the poison-oracle's judgment but will maintain that he is ignorant of his witch abilities. Since accusations of witchcraft are occasioned only by the prior occurrence of some specific misfortune, an accusation always carries with it the claim that the accused used his witchcraft to engineer said misfortune. An accused individual may accept the poison-oracle's judgment, but will deny that they consciously attempted to bring about the misfortune they are accused of having instigated. Evans-Pritchard was unable to find any metatheory espoused by the Azande which would account for or relieve this tension. In fact, further investigation obscured, rather than clarified, the situation as far as Evans-Pritchard was concerned; certain ritual behaviours and social mannerisms he recorded seemed to suggest that it is entirely possible to be a witch, or have witchcraft substance in one's belly, without being conscious of it, apparently contradicting explicit Zande witchcraft doctrine. In the end, he concludes:

Zande doctrines [regarding witchcraft] are so numerous, varied, and plastic, that a man can always find in them an element to serve his interests in any given situation. He does not deny the doctrines, but he selects from them what is most to his advantage in each situation and excludes the rest.¹¹

Evans-Pritchard was later taken to task by Winch for the judgments he pronounced upon the Zande forms of life; we do not need, here, to join that debate. What

¹¹ Evans-Pritchard, E.E., *Witchcraft Oracles and Magic Among the Azande*, Oxford (The Clarendon Press), 1937, p. 133.

is of importance for our purposes is that Evans-Pritchard's observations of the Azande attest to the possibility that a cognitive agent may employ certain ideas and accept them together without making any attempt to systematize them into a coherent body.¹² Where there is no attempt, desire, or motivation to systematize a certain set of ideas, there can be no subjective apprehension of conflict among those ideas; such ideas do not *challenge* one another. Where there is no apprehension of conflict, there is no obvious motivation to revise or reject. Hence, Zande doctrine on witchcraft embodies a form of dogmatism; but this dogmatism does not necessarily arise from *belief* in the doctrine. Rather, it may arise purely because there is no frame of mind that can be entered by Zande individuals which would encourage them to think that any part of the doctrine is wrong.¹³

In other domains, Zande cognition seems to respect many of the desiderata of systematized thought which we generally associate with rational thinking. Hence the Azande may be thought of as cognitive agents with partial systematization of their individual corpora of ideas. It is more difficult to find such clear examples of unsystematized ideas in western culture, or at least any example which has the pervasiveness and persistence of Zande witchcraft doctrine; this difficulty may be attributed to the *general* or even *universal* lack of tolerance which western culture has for inconsistency or incoherence. Nonetheless, the presence of such lack of systematicity is a live option for any cognitive agent.

¹² I am excluding here the possibility that Zande doctrine is in fact systematized, but that Evans-Pritchard was unable to uncover the necessary evidence which would reveal this. It is generally agreed that the Azande do not have what would be judged by Western standards as a coherent body of doctrine with respect to witchcraft.

¹³ Or even apprehend their doctrine *as a doctrine*. Cf. Robin Horton's essay "African Thought and Western Science", from *Rationality*, Bryan R. Wilson (ed.), Blackwell 1970, p. 159.

The sort of cognitive situation which we have discussed in reference to the Zande way of life should not be elided with a superficially similar situation, which may particularly (although not exclusively) affect those educated in the culture of the West. I have in mind the failure to feel any sense of conflict among propositions made across certain domains of discourse – in particular, the domains of common sense and science. The perception of a physical object as coloured, for example, might be imagined as providing a case of potential conflict for a scientifically inclined mind which has accepted the notion that physical objects possess no colour. My present perception of the wooden back of a chair as having a reddish colour might be thought to conflict with my acceptance that physical objects possess no colour, but only properties of spectral surface reflectance. The concern over the lack of friction between these two incompatible ideas is not solved by suggesting that the conceptual framework undergirding my perception of the chair as being composed of a reddish wood is in some way influenced by or ultimately cashed out in terms of notions of surface-reflectance and cellulose fibres and so on; for this suggestion would be simply false. My apprehension of a wooden chair possessed of a reddish colour is not internally translated, in the process of being apprehended, into the scientific notions just mentioned. In the normal course of events, perceptions of this sort are unaltered by exposure to scientific ideas about the nature of matter. The question remains, then, why such divergent accounts do not occasion episodes of awareness of conflict within a cognitive system which possesses them both.

Although there may be merit in the pursuit of the ways in which one domain of discourse may be reduced to another, in at least some cases this may prove to be impossible. The circumscription of certain ways of speaking – and the concepts which

underlie them – into *domains* may, ultimately, be an implicit indicator of an inability to make (or the real impossibility of making) such relations or reductions. In the case of the conceptual domains of common sense and science, it is not implausible to suggest that the relations some of us may make between them will be merely stipulative relations of signification. We may take, in the context of a scientific experiment, for example, a phenomenal experience of reddening litmus paper to be (standardly) an indicator for the presence of acid in the liquid being tested; or take a perception of a reddish wood as a sign of the presence of certain spectral-reflectance properties in the surface of a solid object we observe. In stipulating such relations of signification, however, we do not thereby draw both conceptual domains into conversation with each other; the concepts and the propositions composed within each domain are not of necessity answerable to those which belong to the other. Within each domain, however, conflicts may arise and be recognized. Evans-Pritchard's dialogues with Azande tribesmen, with respect to the inconsistencies of doctrine which he attempted to explain to them, were conducted within the confines of the Zande conceptual domain and the idioms appropriate to that domain. Hence the particular sort of lack of systematicity which Evans-Pritchard apparently uncovered is quite unrelated to the sort of trans-domain non-systematicity just described.

Notwithstanding our taking note of this other form of non-systematicity, it should be mentioned that the extent to which it may manifest itself in any individual depends upon that individual (hence the qualification 'some of us' in the paragraph above). Some individuals may indeed feel a tension between their perception of reddish wood and their scientific conception of physical objects in general; such individuals are likely *not* going to be satisfied with the solution just suggested regarding the relationship between the

domains of science and common sense. In such cases, we must conclude that the individuals concerned with this tension are possessed of a more ambitious tendency to systematize than those unaffected by the apparent incompatibility of certain domains of discourse. Whether this means that they should be further interpreted as having a different understanding of the relationship between those domains or merely suffer from some deeply rooted confusion, we cannot say.

In the analogies drawn above between ideas and genes as both being subject to an evolutionary dynamic, there is an assumption that both the *change* and *persistence* of ideas depend upon both the ability to recognize conflict among ideas and the ability to recognize that a new idea may present a challenge to prior persisting ideas. It is argued, in this essay, that the recognition of challenge and conflict may be frustrated on traditional understandings of cognitive systems that are founded in belief. In the foregoing sections of this chapter, we have attempted to construct a model of cognition which escapes those frustrations. However, we have now encountered a second way in which the recognition of conflict may be frustrated; a cognitive agent may, within a certain domain of ideas, have no motivation to systematize their thinking. Without that internal motivation to systematize – without a *will to systematicity*,¹⁴ to coin a phrase – no rational change of ideas can be expected to take place.

For this second problem, no definitive solution may be offered. We might imagine for consideration some possibilities: that there are certain contexts of cognition where a cognitive system is hard-wired to seek systematization among the ideas participating in that context; or that the desire for systematicity is part of the process of socialization and

¹⁴ Although such a phrase may be literally true of human cognitive systems, its use here is slightly tongue-in-cheek. Such a motivation could take the form of a rule or instruction set in some cognitive systems; it is not intended that possession of a will is a necessary feature of cognition in general.

cultural indoctrination to which a cognitive system is subject. Both suggestions could allow for partially systematized cognitive systems, and there are doubtless other tales that could be told about how it is that the motivation to systematize comes to be part of the operation of a cognitive system. In this essay, we started from the position that such a motivation is present, and that the main problem was to determine how it is possible that conflict could be recognized at all. We have done what we can with respect to that particular task, and will leave the examination of this latter problem, as discussed above, as a separate project.

There is, however, a deeper – albeit somewhat related - problem to consider. We are concerned with the capacity of a cognitive system to recognize conflict among its ideas, when that system is not grounded in belief. A conflict is a *problematic* incompatibility. The notion of problematic incompatibility, as it is here applied, is rooted in the idea of non-contradiction. A standard formulation of the law of non-contradiction runs as follows:

A proposition p cannot be both true and false (at the same time and in the same respect).

The formulation as written is derived from Aristotle. There are others; apart from the typical rendering in symbolic logic¹⁵, however, all versions of the principle make reference to truth or falsity. This presents a potential worry, given our attempt to describe a cognitive system which is not grounded in the truth-ascribing attitude of belief. We wish to consider such a cognitive system as remaining sensitive to conflict among the ideas it has accepted. We must explain, then, how it is that conflict could be recognized – that is, how a problematic incompatibility could be recognized - by such a system, given

¹⁵ For example, $\sim (p \cdot \sim p)$.

that the system is not attempting to adopt a truth-taking attitude towards either of the parties to the conflict. Since there is no attempt to regard either as true, the law of non-contradiction as stated above seems to play no role in the recognition of incompatibility as problematic.

Note that our problem here is not the recognition of incompatibility *alone*. A cognitive system which is in a position to compare two propositions, p and $\sim p$, can recognize their incompatibility within classical binary logic. For the incompatibility to be problematic for such a system, it would seem to follow that such a system must take classical binary logic (or any variant logic which recognizes the conjunction of p and $\sim p$ as false) seriously. It would appear to follow further that such a cognitive system must also take the law of non-contradiction seriously. If this is the case, we may well wonder how it is that said system can genuinely be regarded as treating the contents of its consciousness as merely accepted rather than believed. To answer this question, we might ask why any cognitive system would take seriously at least one of the family of logics that treat as false propositions of the form $(p \cdot \sim p)$.

Michael Dummett once remarked that it is part of the concept of truth that we aim at making true statements.¹⁶ We might take issue with the formulation but it is difficult to find fault with the spirit of his claim. His comparison in the same paper of *truth* and *falsity* with *winning* and *losing* as, respectively, features of a theory of logic and of chess is apt. We are interested in a class of cognitive systems which are truth-oriented; these systems treat the attainment of truth as a goal, and as a regulative ideal for their operations. The collection of falsehoods is to be avoided as much as possible. The tools that such systems have at their disposal for making such determinations are, notoriously,

¹⁶ Dummett, Michael, "Truth (1959)", from *Truth and Other Enigmas*, Gerald Duckworth & Co. Ltd. 1992.

extremely limited and possibly inadequate for the task. Those limitations certainly interfere with the capacity of such systems to come to have *knowledge* of whether their ideas are true or false; yet those ideas may be nonetheless either true or false. If truth is considered a goal, if it is understood that truth and falsehood are mutually exclusive conditions for any individual proposition, and if there are certain conditions under which a cognitive system may recognize a proposition as necessarily false, then that system is bound to reject such a proposition.

Let us return to consideration of conflict among accepted ideas within a cognitive system which belongs to the family of systems just described. We shall consider two ideas with propositional form, p and $\sim p$, which the cognitive system has come to consciously accept. The system does not have knowledge of whether *either* p or $\sim p$ is true or false. One or the other may be true; however, it cannot be that both are true. At least one *must* be false. As the system treats of truth as an ultimate goal and as a regulative ideal, the *de facto* conjunction of the proposition and its negation (for both to be accepted is for them to be conjoined) is unacceptable; hence the persistence of both propositions within the system is recognized not only as constituting an incompatibility under the rules of classical binary logic, but a *problematic* incompatibility. One or the other proposition must be rejected. Insofar as the cognitive system respects the regulative ideal of truth, it is compelled to implement a decision procedure to determine which of the two, if either, may be retained as acceptable. The decision procedure used need not (and probably cannot) yield a determination that one or the other of the incompatible pair is *in fact* true (or false). For any proposition that is not self-negating, both acceptance and rejection are always provisional; the immediate goal of the decision procedure

implemented in order to eliminate the conflict is to choose one over the other based on some criterion or set of criteria, such criteria being themselves open to revision.

It may be that the most a cognitive system can hope for is to possess as accepted a body of propositions all of which are true, without ever coming to know - in the classical sense of the word - what their alethic status is. There is no reason for this unfortunate possibility to deter us from regarding the acquisition of truth as a worthy goal; to that end, we do ourselves only favours by performing as much cognitive housekeeping as is practical.

II.D. Acceptance, Interpretation, Belief

The family of cognitive systems discussed in section II.B suggests certain qualities that the interpretative mechanism required by these systems must possess. Such a mechanism must produce, for consumption by such systems, a single interpretation at any one time; because there is no inspection stage, we cannot allow an interpretative mechanism which produces multiple interpretations simultaneously from any given experience. The interpretation of new experience offered will be, in any particular case, either of a sort which assimilates easily into the set of ideas already accepted by the corpus, or of a sort which actively conflicts (and is recognized by the system as conflicting) with some subset of accepted ideas within that corpus. As we discussed in the chapter which first outlined the nature of the problem, a perceptual interpreter which includes an inspection stage may be influenced in its operations through constraints imposed by what is already believed. This influence, as we have seen, leads to an

inability either to construct or to approve interpretations of experience which do not accord with what is already believed. It may have occurred to the reader that a similar problem besets our attempt to describe a cognitive system grounded only in acceptance. We should be concerned that the very argument brought against the theory of cognition grounded in belief – that perception is detrimentally influenced by prior belief – may also be brought against a theory of cognition grounded in acceptance. If we were able to describe an interpretative procedure which required no inspection-stage and which at the same time could yield perceptions which were at variance with what was previously believed – in other words, could yield perceptual episodes which were not constrained in their interpretation by prior beliefs – then we could dissolve the problem and accept that a cognitive system grounded in beliefs may well describe human cognition. It might seem that, were we able to describe such an interpretative mechanism for a cognitive system, the notion of acceptance would then be superfluous. We shall now attempt to describe such an interpretative mechanism, and then further attempt to show that, notwithstanding any intuitions which suggest otherwise, such a mechanism could not be part of a cognitive system grounded in belief.

To begin with, any such mechanism must first escape the binding power of what has previously been accepted. Yet neither can the mechanism be completely unbound, able thereby to generate interpretations based on anything that the individual may conceive. There must be some middle way. Let us investigate the notion of an interpretative mechanism which is grounded in a form of *association*.

What may be suggested is that this interpretative engine has access to a library of concepts that have nothing to do with what the cognitive state of the system currently

regards as constituting its local environment. The library must be at least somewhat atomic in nature;¹⁷ the interpreter may be described as having built, through habit and experience, *associative* pathways between particular data points available for extraction from bottom-up processed information packets, and the concepts (or concatenations of concepts) in the library. The associative pathways must be capable of being both many-to-one, and one-to-many. Many-to-one, because many different bottom-up deliverances of the sensory system may be associated with the single complex concept of (for example) a *cup*. One-to-many, because any single bottom-up deliverance may be associated with any number of complex concepts. Any particular association which may be made between a bottom-up processed packet of information and a top-down application of a complex concept is underdetermined; nothing in the interpretative engine mandates that one particular associative path is *correct*, and all are ultimately *acceptable* associations as far as the cognitive system itself is concerned.

The criterion which determines which associative pathway will be followed in the production of a particular interpretation is the *quantity of processing power* which may be devoted to the interpretation of each element in the information packet in question. The more processing power available, the more data points may be considered in the production of an interpretation of experience. A difference in the number of data points in the packet which are used for generating an interpretation may alter the interpretation generated, by altering the associative pathway from one to another of the group of pathways that may be linked to the information packet.

¹⁷ Moreover, the library may even be subdivided into different collections corresponding, among other things, to the various sense-modalities available to the system.

It is reasonable to assume that in animal cognitive systems, relatively few data points are used in the generation of interpretations of any particular information packet; cognitive resources are limited and in most situations economies must be made. Generally, the more data points that are attended to, the more detailed the perception: additional information will merely *refine* what is perceived. In the act of jaywalking, for example, most available processing power is diverted to judging the distance and speed of the automobiles approaching one's current location; although one may apprehend in one's visual field an automobile, one will not attend to the make and model, nor whether the upholstery is leather or cloth, nor the number of passengers – even though all of this information may be present to be extracted from the image apprehended. However, merely standing by the side of the road and apprehending the same image may allow the viewer to remark upon all of these details as more processing power is available for the task. More importantly for our purposes, attending to more data points may also *shift* what is perceived altogether. Some trompe-l'oeil illusions, for example, trade on the casual way in which we generally attend to information made available to us in experience. Pure night-vision experience also provides a dramatic example of the way in which interpretations of visual experience may shift, sometimes oscillating in a cycle partially influenced by visual artifacts caused by retinal or other ocular imperfections which are more pronounced in low-light conditions. In daylight conditions, the interpretation of objects perceived in peripheral vision can sometimes radically alter when the center of visual attention is shifted towards them.

Such a mechanism shares some features with one of the mechanisms discussed in that chapter where the problem was first described. There we proposed a mechanism

which possessed no inspection stage and in which the interpretative engine was not constrained by what was already believed. Two subvarieties were then described: one in which interpretation of a single visual scene varied from moment to moment as the interpreter generated interpretations at random; another in which the available interpretations were superimposed upon one another. In both subvarieties, the output of the interpretative engine was considered to be believed. As a result, both were rejected as possible characterizations of human perceptual processes as the conscious apprehension of perceptions generated by such mechanisms do not fit well with our own experience.

The proposal just described is similarly unconstrained by prior acceptance, and also requires no inspection-stage; its advantage over these earlier attempts lies in its ability to deliver a single interpretation at the termination of any particular episode of interpretation, and for that interpretation to remain stable through subsequent momentary apprehensions of similar raw phenomenal data. However, because the interpretation generated for delivery to the conscious portion of the cognitive system may be altered simply by a variation in the amount of processing power devoted to its generation, it seems unintuitive to think of its deliverances as beliefs; they represent only a possible interpretation of a state of affairs. Nonetheless the interpretation is delivered as *accepted*, as it represents the outcome of one of the many acceptable associative pathways which generated it. In many cases, the interpretation delivered will cause no friction when the cognitive system attempts to integrate it into its current state of understanding of its local environment, and so will be simply absorbed without question. In case of conflict, we might expect that the first reaction of the cognitive system will be the diversion of more processing resources to the *reprocessing* of the raw information which led to the

interpretation.¹⁸ This reprocessing may or may not alter the associative pathways followed between the bottom-up processed information given to the top-down interpreter, and hence alter the concept(s) which, when applied to that information, complete the interpretation and generate a real perception. If it does not, then it may merely issue in a more refined version of the same conflicting interpretation that initiated the reprocessing. In any case, the result will be delivered, once again, as accepted. This process may repeat until one of the following occurs: (a) the greater concentration of processing power yields an interpretation which does not conflict with what is already accepted; (b) The maximal number of data points possible is extracted from the raw information and the resulting interpretation is delivered. If this most refined possible interpretation still conflicts with what has been previously accepted, other processes come into play to decide what may be kept and what discarded or otherwise revised.

The general schema is simple enough. Nonetheless it is still possible to object that the interpretations yielded by such an associative framework could be believed, rather than merely accepted. The intuition which guides this objection is not entirely without foundation; there is an aspect of each perception, even within this model, which may be susceptible to being described as believed. I have in mind a second-order apprehension, not of the perception itself, but of the phenomenal fact of its apprehension. Articulating this with a concrete example, one might say that, of the perception of a piece of white litmus paper a portion of which has turned red, we may speak of being ‘appeared to whitely’, ‘appeared to redly’, and ‘appeared to litmus-paperly’, among other features. These second-order apprehensions would be fair candidates for aspects of a particular perception which could be objects for a truth-taking attitude. For human cognitive

¹⁸ If not reprocessing, then processing of the next available stimulus within the same context.

systems, it seems particularly natural to say that such apprehensions will be designated as beliefs in the reflective mode of thought which characterizes the cognitive context of decision.¹⁹ However, among the features of any particular perception itself, nothing is naturally susceptible to being so regarded.

Given the repeated emphasis throughout this work on the division in the interpretative processes of perception between bottom-up and top-down processing, it might have appeared to the reader that an argument could be made with respect to the bottom-up portion of interpretation, such that the features of a perception contributed by that portion could themselves be taken as believed; indeed, when we were discussing belief-based cognitive systems in the exposition of the central problem, we even allowed that such basic processes might automatically issue in rudimentary beliefs to which top-down processing might be applied. However, the sort of cognitive system now being discussed is not so grounded in belief, and so there is no reason to think that the bottom-up processes of interpretation need issue in any epistemic evaluation at all. On the present model of cognition, application of the attitude of belief or any other epistemic evaluative attitude is an event that may only take place in the context of decision, after the fact of the formation of the perception in question. And yet, because of the description of bottom-up processing as given – that being that its mechanics follow a fixed procedure and its deliverances are not open to any sort of revision or adjustment despite being conceptualized – one might continue to suspect that a cognitive system examining a

¹⁹ Doubts may arise even here, by appealing to the fallibility of memory; since the examination of a perception for the purposes of evaluating it or its features occurs after the fact of its acceptance, one *might* try to claim that one was not, for example, really appeared to redly. So one might have to fall back on a mere general claim concerning the perception, such that one could maintain that one had the perception one had, *in just the way that one had it*, and no other; and then take a truth-regarding attitude towards that claim.

perception after the fact of its acceptance ought to designate these unrevisable features as believed. In order to address this potential concern, we might describe two different sorts of unrevisability for cognitive systems possessed of bottom-up interpretative mechanisms. On the one hand, there is the unrevisability which may be said to be a feature of the second-order apprehensions described above: one had the perception one had, in the way that one had it, and no other way. It is extremely difficult to imagine ways in which this could be denied. On the other hand, there is the unrevisability which may be said to be a feature of the deliverances of bottom-up processing: it may be a matter of fact, for example, that visual perceptions will present such features as colour, depth, distance and even motion to human cognitive systems, and that we have literally no control over the inclusion of such features in our visual perceptions; these features are beyond the reach of interpretative adjustment for us. But this does not entail that we must evaluate such features as believable when we come to examine them in the context of decision. For example, one may be looking at a trompe-l'oeil of a red ball painted onto a wall. It may be, in human perception, simply impossible *not* to apprehend that experience as being of a red, round object, that is, impossible to apprehend that experience in any other way, notwithstanding one's awareness, perhaps for independent reasons, that what one perceives is an illusion. When standing in the appropriate position for the success of the illusion, one may be conscious of the illusion and yet unable to modify the presentation of the perception as being of a red, round object. We would not, in this case, be in any way compelled to take it as true that there is a red, round object before us, notwithstanding our total lack of control over the bottom-up conceptual matrix which yields such features of perception. Hence, although it is plausibly natural to apply a truth-

taking attitude towards *certain* aspects of a particular perception in the context of decision, the scope of such application is extremely limited.

In the act of perception itself, the interpreter itself has no control over how much processing power may be devoted to any one interpretative episode, and so may yield entirely different interpretations from the same information packet depending upon how much power the cognitive system is able to spare for its labours. With respect to each preprocessed information packet, each potential interpretative pathway is, at any given time, an acceptable pathway for generating an interpretation; and so too is each interpretation which may be produced by following any of these paths. The sole criterion (as mentioned above) during any interpretative episode, for the actual production of one or another interpretation, is the quantity of processing power which is devoted to the task. Such a criterion does not fit well with any sort of *logical* argument that may be made concerning the legitimacy of the interpretation delivered. Since the variability of interpretation is essentially controlled by a form of *willful ignorance* rather than simple unavailability of data, it is difficult to argue that any particular interpretation constitutes, for example, a 'best explanation' form of induction. The system must recognize that, given more time and more resources, the interpretation finally delivered by the interpretative process might have been different from the one actually delivered; and so it seems an unintuitive move to suggest that the system will take a truth-regarding attitude towards any interpretation actually delivered as a perception.

Yet, while the model system may indeed be coherently described by making use of the proposed notion of acceptance alone, and while it may even be allowed that it makes a good deal of sense to do so, proponents of belief-based cognition may yet

register an objection. Suppose that, notwithstanding the argument, we nonetheless tried to describe the operations of the model system in terms of belief. The deliverances of the proposed perceptual processor as described may well vary depending upon the quantity of processing power devoted to the mechanics of interpretation; yet would there be any deleterious consequences upon the model as a whole, if we were to consider these deliverances as beliefs? There is no obvious reason to suppose this; and if there are no such consequences to be derived from so doing, then there is no longer any strength to the argument that being able to revise one's beliefs is impossible, even in the limited although crucial way that the discussion of the problem has attempted to show. Here is an example: I intend to jaywalk. I glance up the road at oncoming traffic; I attend primarily to the speed and distance of the approaching automobiles, but also manage to perceive that the closest car is occupied by two persons. Now suppose I register this perception as a belief: I believe that the car is occupied by two persons, notwithstanding that I used only a very small portion of the available quantity of processing power in the production of this interpretation of the local environment (I also now believe that the cars approaching are doing so at a certain speed and that they are at a certain distance from my current location, and have devoted most of my available processing resources to making these determinations). I now look across the roadway and judge the distance as too great to make it across safely, given what I have come to conclude concerning the speed and distance of oncoming traffic. So I decide to stay put. I then look back at the oncoming traffic, now perceive that the closest car is occupied by one person only, and take this perception onboard as a belief. This conflicts with my earlier belief concerning the level of occupancy of the vehicle. The conflict occasions reassessment of both the

earlier and later beliefs; and, because the later belief is the more studied of the two, having been arrived at through more thorough attention to a greater quantity of data than the first, I reject my earlier perception as false and attribute its origin to the small quantity of attention devoted to its production. Here is an example of belief-based cognition which operates within the constraints of the proposed model, is subject to none of the worries of the earlier descriptions of the central problem, and allows me to revise my beliefs.

Therefore while it must be admitted that, if belief-based cognition is what is wanted then one must adhere to some description of cognition which includes the appropriate sort of independence for the perceptual processing mechanisms (the model proposed here being a good example of the sort), such models of cognition do not in and of themselves motivate exclusion of the possibility that belief-based cognition can suit as descriptive of human cognition or cognition in general.

It is an excellent objection; yet it does not entirely defeat the argument concerning the impossibility of revision. What it does suggest is a further restriction of the contexts in which revision of belief may be impossible on a belief-based construal of the model. What follows is an example that illustrates this further restriction. Let us suppose that the objection is accepted and the proposed model for cognition is treatable as belief-based. Suppose I perceive a black bird of a certain shape and size, and with other distinguishing characteristics. Since the model is now being treated as belief-based, the perception is a belief. Now suppose I perceive another black bird possessed of characteristics remarkably similar to the first; and again, yet another such bird. Now let us suppose I perform an enumerative induction of a familiar sort, and conclude that all such birds which conform to the set of characteristics observed excluding the colour, are black. For convenience, I

shall give the name 'raven' to my conception of birds which conform to such characteristics (again, excluding colour). This general hypothesis can be rendered 'all ravens are black', and this proposition is a belief. I can allow that in some possible world, there are white ravens, and that the belief that all ravens are black may be falsified if I happen to be presented with a non-black raven. Now let us recall that on the model currently being proposed, the perceptual processor operates independently of the main body of extant belief. In order for that processor to perform its task, the associative pathways between bottom-up feature processing and the top-down application of more sophisticated concepts will be subject to revision and addition over time, to accommodate what the cognitive system has learned. If we now make use of the newly formed general hypothesis to build a new associative pathway, it might take something like the following form. A pathway will be built such that: given the bottom-up deliverance of a particular sort of shape, a separate bottom-up deliverance of a certain visual texture, and a bottom-up deliverance of black coloration corresponding to the area that this shape occupies in the visual field, the conjoined presence of these processed features will activate an association with the concept *raven*, and that concept will be duly applied in the formation of future perceptions. The concept *raven* is, as noted above, a complex concept which comprises – among other features – a concatenation of certain potential bottom-up conceptual deliverances including a certain shape and certain visual textures, but not any particular colour. The formation of the general hypothesis concerning the coloration of ravens need not entail the alteration of the concept *raven* such that it comes to include the colour black, thereby making it analytic that ravens are black. However, it is reasonable to suppose that the cognitive system will adjust the operations of the perceptual processor

to accommodate this new, general belief; this adjustment will take the form of construction of an associative pathway that only allows application of the concept *raven* to be triggered when the shape and other visual features are experienced in conjunction with the colour black. We could conceive of the associative pathways of this model as being handled or brokered by the device, familiar from the discussion in cognitive science of parallel processing models for artificial intelligence, of a weighted node; adjustment of the associative pathways within the processor or the development of new pathways would consist in adjusting the weights on such nodes. Then we might say that the absence of input to the node of a bottom-up deliverance of the colour black, even supposing that the other required bottom-up features are present, would not place enough signal strength on the node to open the pathway to the concept *raven*. Hence perception as of an object of the appropriate characteristics, but of a colour other than black, will not trigger conceptualization of the object as an instance of *raven*; the final perception delivered by the perceptual processor, however it turns out, will not be rendered as a perception of a non-black raven. As a result, no future perceptual belief will be formed which could in any way undermine or challenge the general hypothesis that all ravens are black. This returns us to the central problem of earlier chapters: viz., that extant beliefs will thwart interpretations of experience that might conflict with those beliefs.

Such a situation would not arise if the same process were described using only the notion of acceptance. Having formed a general hypothesis concerning the coloration of ravens, but not taking that hypothesis as true, would allow the cognitive system to form a somewhat different associative pathway between the relevant bottom-up features and the concept *raven*. The system might build a node, or nodes, of the appropriate sort in the

perceptual processor to allow that the presence of the relevant bottom-up features in conjunction with the colour black would trigger application of the concept *raven*; yet because it is not *true* for the system that all ravens are black, there would be no reason for it to so weight the node(s) as to prevent application of the concept *raven* if the relevant bottom-up features were to be presented in conjunction with (for example) the colour white. The general hypothesis is accepted by the system and may play a role in determining action and prediction; but as it is not taken to be true, it is unreasonable to suppose that such a general hypothesis might be used to rule out certain associative pathways in the perceptual processor.

The obvious response from the objector might be to suggest that a belief-based cognitive system which is sufficiently conscious of its own fallibility would behave in the same way as an acceptance-based cognitive system, and so would not cut itself off from the potential for perceiving a non-black raven through the construction of the sort of limiting associative pathway as was suggested above. However, given the unresolved tension surrounding what was earlier described as common fallibilism and its propensity to collapse into strict fallibilism, acceptance of the objector's response requires argument; and there is a very real concern that provision of more argument to support the response will undermine the plausibility of regarding the (now explicitly fallibilistically construed) belief-based system as having a truth-regarding attitude towards its own thoughts, and in particular, the thought that all ravens are black.

Even on the assumption that this reintroduction of the tension in common fallibilism may be resolved, accepting the response and conceding that it allows a belief-based model ultimately to avoid the central problem has a moderately unhappy

consequence. It would be of great value for the model proposed as a solution to the problem to have some chance of being descriptively correct not only of the actively critical and reflective mind, but also of the uncritical and unreflective mind. The final response, however, requires that the cognitive system explicitly adopt a fallibilistic attitude towards its own cognitive processes and judgments, and it is not plausible to assume that this is a natural feature of an uncritical mind. Hence we would then be able to escape the central problem while holding on to the notion that belief is a central feature of cognition only at the expense of adopting a model which had little chance of being descriptively correct, both of a significant number of adult human minds (though we might wish it were otherwise), and very plausibly of all juvenile or developing minds.

Conclusion

The exploration undertaken here has led us in some peculiar directions, and there is no shortage of dead ends. If chasing these red herrings became tedious for the reader, I can express sympathy, but no regret for having included them; for such pursuit, it seems to me, helps to illuminate the paths we might wish to take.

If the arguments in the foregoing pages can be judged internally consistent, there remains the question of what value these lucubrations possess. It would be most unfortunate, if the exploration of a problem I had originally thought of great significance turned out to be no more than an academic debate concerning the use of a technical term in the discourse of epistemology (viz., belief). Certainly the articulation of the problem and the subsequent attempt both to explore its tenacity and determine its structure may be partially judged an exercise in terminological disambiguation. As such, it will be of direct value almost exclusively to professional philosophers practicing in the analytic and post-analytic tradition. The debates in epistemology as practiced within that tradition may be enriched by its articulation, or not; no one else will have much to say about it, despite the fact that it may be applied as much to the common-sense¹ understanding of the concept of *belief* as to the academic philosophical sense.

In the introduction, I remarked that the investigation of the issues at hand was conducted in a Humean spirit; but the early stages of the discussion clearly owe

¹ 'Common sense' as understood as the common sense of western culture, which is perhaps uniquely infected with the lack of tolerance for contradiction and incoherence characteristic of the scientific attitude.

something also to the spirit of Ryle, that great hunter of conceptual confusion. While the confusion I initially hunted was one which I thought bound up in the concept of *belief*, it inevitably also became a hunt for my own conceptual confusions. In the early stages of the project, my first poorly articulated thoughts concerning the central problem led me to think that perhaps belief had no role whatever to play in rational human cognition. In the course of developing the ideas contained in this essay and answering objections to same, I came to realize that even though the central problem may be real, the possible solutions to it comprised a more complex terrain than I had at first thought. (These observations concerning my own naïveté likely come as no surprise to the reader.) Hence I came to abandon my early idea that we do, as a matter of fact, get along wholly without belief; and yet, given the central problem, the various ways of describing human cognition which I had encountered in my studies left me wholly unsatisfied. It struck me then that there must be some description which would avoid all the difficulties which the problem raised, without forcing me to adopt any presuppositions which grated on my existing skeptical intuitions – while at the same time respecting certain constraints which may be tied to the phenomenology of conscious human thought and experience, including the felt passivity of perception as opposed to cognition, and the ability to alter one's opinions rationally without requiring the possession of either a skeptical outlook, or some form of common fallibilist perspective. The positive account I offer in the latter stages of the essay, I think, finds or at least points towards such a description.

The positive solution to the problem offered – the construction of a model cognitive system which can accommodate said problem – may possess an intellectual appeal independent of the problem which led to it. Although the development of the

model was guided by a sensitivity to certain skeptical concerns, the model itself does not depend on such concerns for its operation. The model does not seem to – and is certainly not intended to – entail that cognitive systems of this type possess some sort of appreciation for the sorts of skeptical problems that arise, paradigmatically, from reflection upon the Cartesian method; nor, indeed, does it require that the system operate under any sort of epistemological sophistication of the sort which leads to fallibilism, or an appreciation of the distinction between talk of appearances and talk of what is the case.

If this is correct, it would be an important feature of such a model insofar as it may be used to describe human cognitive systems of varying degrees of maturity and sophistication. It is unlikely that we are born with appreciation of skeptical concerns of any kind; such philosophically sophisticated worries may only come to us in the course of our intellectual development, and then none too early in that development. Those who come to appreciate skeptical reasoning will likely also come to appreciate, in some fashion, the odd lack of influence such reasoning seems to have over cognition as it is carried out in ordinary circumstances, and may feel that this peculiar feature of our thinking needs explanation; surely, many philosophers are in this predicament. The acceptance-based model sketched here may provide such an explanation. For, the adoption of an attitude of doubt towards an idea (or more properly, the proposition embodied by an idea) is as much a form of evaluation as the adoption of a truth-taking attitude, and if cognition ordinarily proceeds without reference to the evaluations that may or may not be attached to our ideas, then an individual's skeptical concerns need not intrude upon their game of backgammon. Hence this model may be offered as a possible

description of a cognitive system which is beset by skeptical concerns, and in so doing, may describe a form of livable skepticism; this is one way in which it may be of philosophical interest. Of especial value, however, is that the model may also plausibly be applied as a description of an undeveloped or juvenile mind, or indeed any mind in which skeptical considerations have not arisen – for surely there are many adult minds which remain untroubled by skepticism of any depth. Such underdeveloped or untroubled minds seem quite capable of revising their opinions in spite of the presumed lack of philosophical sophistication of any sort. The model cognitive system proposed here may be used to describe cognition for these sorts of minds as well.

Let us look at the essential elements of that system. We proposed that it have, on the one hand, an interpretative mechanism for processing raw phenomenal experience, and on the other, a repository for all such processed information, which repository holds in addition any other information that the system may contain. The interpreter itself comprises a complex of processes. Initial-stage processing of raw information corresponds to what is generally referred to as ‘bottom-up’ processing: the imposition of conceptualized structure over which the system has no control.² In processing visual sensation, this may include such information as colour, dimension, distinctions of foreground from background, and so on. This partially-processed information is then further refined by what we provisionally have labeled ‘top-down’ processing; however, our conception of how top-down processing is implemented is somewhat different from more traditional proposals. Rather than the active participation of extant beliefs, this part of the interpretative process requires a set of structures whose interpretational dispositions, as it were, are configurable with content which may be derived from extant notions which

² This would be true of human intelligence, at least.

inhabit the system's main repository – yet which structures are in fact logically separate from such a repository and operate independently of it. This crucial separation allows that there be a part of the process of interpretation which is at once influenced by how the system conceives of itself and its environment, and yet is not unduly sensitive to the system's current state of its conception of itself and that environment. Rather, the configuration of the top-down interpreter operates at one remove from that current state, and can be configured with a set of parameters which may be said to correspond to how the deliverances of the bottom-up portions of the interpretative process are *generally* interpreted. This provides a first step towards a way for a cognitive model based on acceptance to avoid the pitfalls of overinterference by extant thoughts on the process of interpreting new perceptual experience, which plagues belief-based models. However, more is required, both in order to dispel our concern more thoroughly, and also to explain why the model does not fit well with a belief-based system. For the parameters of this interpretative processor cannot be set so as to mirror any extant, fully-formulated thoughts which the system may possess; rather, it must be configured to be responsive only to the conceptual framework within which that system's thoughts are constituted, with bias being assigned to certain sorts of interpretation based on prior experience and deliberation. It is denied that the *current* state of the cognitive contents of the system could play any significant role in influencing the *current* configuration-state of the processor, as the adjustment of bias is a slow and complex procedure: such adjustment is, it might be said, *diachronic* rather than *synchronic*. The artificial-intelligence model of weighted nodes for the processing of input to output comes to mind as a possible instantiation of such a mechanism. Once this is allowed, it opens the door to

consideration of how individual concepts or complexes of concepts may come to figure in the interpretative process. Here it was proposed that the application of particular conceptual resources and the associated biases in processing, during any particular episode of such processing, will be governed entirely by the *quantity* of processing power which the system decides to devote to such interpretative processing. Because the devotion of limited resources in a finite cognitive system is a matter of choice (although the choice is often enforced by practical considerations), it seems unintuitive to conclude that the final output of such interpretation will be, or ought to be, considered by the cognitive system as believed (as opposed to accepted). In addition, the particular model of perceptual process here described (in virtue of its being diachronically adjustable) may become too rigidly configured to allow the sort of potential for conflict among our thoughts that we desire, if we allow that beliefs continue to play a central role in cognitive processing in general.

The top-down portion of the mechanism as proposed, then, has the following desirable characteristics: it operates at one remove from the extant cognitive state of the system; it is adjustable; it makes use of conceptual resources with certain interpretative biases, both of which may arise and be adjusted through experience and development; its deliverances may vary with the quantity of processing power available; its deliverances may be countenanced as perceptions although they are not beliefs: no truth-taking attitude is adopted towards these products of the mechanism; and it dovetails with the intuition that perception is, by and large, a *passive* process.

The perceptions are delivered by the interpreter to the repository of all information which the system possesses. Whether this repository is itself structured in

accordance with a holistic, Quinean model or a Cherniak-style model is matter upon which we may remain silent; nothing, for our purposes, depends on taking the one or the other – or any one of a number of other structures – as definitive.³ As was discussed in the previous chapter, the contents of this repository may here be subjected to any number of cognitive operations with which we are familiar, including both logical operations and the business of coordinating and integrating information. In the normal course of such operations, no epistemic evaluation need play any role; there need not be any assignment to the individual constituents of this repository of such evaluations as *believed*, *held probable*, *held with justification*, or any of the other sorts of evaluation we consider constitutive of the possession of knowledge. Such evaluations only come into play when they are required, as in cases of internal conflict between individual thoughts (which includes perceptions). They may also be brought in when the system spontaneously decides to consider such issues as the origin of the presence of a particular thought within the system, or more generally some sort of explanation relating to any particular thought. Then various mechanisms of evaluation may be devoted to the closer scrutiny of a particular thought or collection of thoughts. These mechanisms themselves, as manifested in human cognition, are likely a mixture of processes; some of which are innate to the system, and some which are introduced in the course of its development. Those methods and mechanisms which are introduced in the course of maturation are subject, for their use, to the same vestigial license that applies to the thoughts accepted into the system – their mere persistence. Such learned methods invoked in the contexts of decision, conflict resolution, and evaluation may be adjusted, rejected, or reformulated, in much the same

³ Practical considerations clearly favour the Cherniak model but the overall function of the proposed system does not depend upon his description of cognition.

manner and for the same reasons as any of the thoughts in the system: because of conflict with other accepted methods, failure to produce workable results, and so on. Included among these may be some forms of induction and probabilistic reasoning, although we can allow that such methods may in fact be hard-wired.

However, having attempted in this way to dispose of belief – construed as the truth-taking attitude – as being central to a theory of human cognition, we seem to have left no place for it at all. For, even in the contexts of decision and conflict-resolution referred to above, while one might have the *capacity* to assess any particular thought as believed, we argue that this cannot be the case on pain of reintroducing the problem we wish disposed of. Nonetheless, it is highly intuitive to think that we may have had contact with belief, and in particular, belief of a sort that is bound up in some way with perception. Where might we find it?

It may be found primarily in the second-order apprehensions which may be made of a perception, after the fact of its acceptance: it is extraordinarily difficult to deny that one had a particular perception in the way that one had it, and no other. If concerns about the reliability of memory intrude here, then one may want to allow that one is misremembering that one was, for example, “appeared to redly” in the course of a particular perceptual episode; so one might then want to deny that particular such second-order apprehensions are good candidates for belief since they may be open to this sophisticated form of doubt. Nonetheless the general second-order claim to the effect that (speaking of a particular perceptual episode) the perception was had in the way that it was and no other, without specifying the ways in which it is now remembered as having been had, remains very difficult and maybe impossible to deny.

It might have been thought that the division in human perceptual processes into bottom-up and top-down mechanisms also provides a way for belief to get a foothold. The workings of said bottom-up processes are cognitively impenetrable; there is no point, in this part of the process of producing perceptions, over which a human cognitive system has any control: it may only accept what is given. The immediately succeeding top-down portion of the process may be cognitively impenetrable as well, but only to a certain degree. There is no direct control that may be exercised over it in the act of perceiving; however, this part of the process is nonetheless open to adjustment by the system. Because the top-down mechanism is adjustable and depends for its operations upon access to learned conceptual resources, those portions of completed perceptions which owe their generation to the action of this top-down mechanism will be regarded in a different light, should a perception come to be more closely examined in the context of decision. We might say that, in the examination of a particular perception within this context, there is a portion of it which may be *thought differently*, and a portion which may not. Recall the litmus-paper example discussed earlier. The perception delivered by the interpreter in this case may be decomposed into, on the one hand, a conceptualized apprehension of a reddening of a portion of a white object distinguished from other background objects, and on the other a superstructure of further, more refined concepts which categorize that white object as paper impregnated with litmus, saturated with the liquid precipitate to be tested in the experiment, and so on. For human cognitive systems the former collection of conceptualized information is resistant, in the strongest possible terms, to any sort of reinterpretation. The balance is quite open to further interpretation (this whiteness is not really paper; the precipitate *itself* is not reddening; there is no lab,

no paper, no experiment: I am a brain in a vat). Yet despite this division in perception, no motivation may be found for thinking that our inability to imagine concrete alternative interpretations of the deliverances of the bottom-up portion of perceptual processing should thereby make such deliverances susceptible to being judged, in the context of decision and evaluation, as beliefs. The bottom-up portion of (for example) visual perceptual processing, for us, always issues in perception as of coloured objects embedded in three-dimensional space, and in time. While one can conceive that, in reality, spacetime is a twelve-dimensional continuum and that objects (such as they may be) possess no colours, it is not possible for us to perceive in this manner, nor even possible for us to imagine what it might be like to perceive in this way: it is not possible for us to think differently with respect to such bottom-up features in any concrete terms. Our appreciation of the possibility of alternatives in the abstract, however, prevents us from reasonably taking such features as we perceive them as true, when we come to examine them in the contexts of decision and evaluation.

The types of cognitive systems we have here described, however, are nonetheless free to take a truth-regarding attitude towards any of their perceptions or other thoughts, notwithstanding the lack of any reason to do so, and notwithstanding the consequent future inflexibility of perception and cognition that would result. For unreflective, uncritical, or juvenile instances of such systems, we might want to argue that the attitude of belief is more commonly taken at the very least with respect to such bottom-up features of perceptual processing, when these are examined in the contexts of decision and evaluation, by virtue of the lack of awareness in such systems of alternative conceptions which might militate against taking a such an attitude; but there is no strong

argument for suggesting that this must be so. Against this idea it might be argued that it is implausible to suppose that such uncritical systems would engage in such sophisticated dissection of their perceptions at all. As for such uncritical systems taking a truth-regarding attitude towards those features of its perceptions which we have classed as top-down, it might be further argued that the very condition of conflict, which is here supposed as the most general reason for entering the cognitive contexts of decision and evaluation, would itself engender the awareness of alternatives that could help to prevent a cognitive system from taking a truth-regarding attitude towards the winner of any particular contest of this sort. In any case, we must allow that a cognitive system which is free to do what it wishes may always decide, even without reason, to impose such a judgment upon its perceptions, features of its perceptions, or more generally, any of its thoughts. So much the worse for an individual system which decides to do so.⁴

So here is a place for belief; the significance of its role in cognition far less important than generally supposed, but present nonetheless. It is of interest to note that by constraining what may rationally be believed to second-order thoughts about the presentation of perceptions, no such beliefs which an individual cognitive system may possess will themselves conflict with each other; for all such beliefs will relate to individual episodes of perception. Moreover, once the conflict between thoughts which gave rise to the context of decision in which such judgments are made has been resolved,

⁴ It is *possible* that the Azande are prone to such a condition; the tolerance, in their culture, for notions which we might regard as cohering poorly or not at all being no bar to the continuing vigor of that culture. As noted earlier, the Azande do not seem to be in thrall to the systematizing urge which drives the culture of the West; hence apparent contradictions in their doctrines, pointed out by inquisitive anthropologists, do not trouble them or may not even be comprehended as contradictions. As a result, for the Azande to possess *belief* in such doctrines will be of little consequence as they have no intention of maintaining any flexibility of attitude towards those doctrines in any case. We say 'so much the worse' only because acquiescing in *belief* towards such doctrines may turn out to have deleterious consequences should there be a change in their environment with which their normal, doctrine-driven practices cannot cope.

this evaluation – as well as any other evaluations made upon those portions of a perceptual interpretation which are open to revision – will recede into the background, having no role to play in the normal course of cognition: the reception and integration of new information which does not conflict with any thoughts already accepted. Hence the mere presence of these sorts of beliefs within the system proposed is unproblematic.

Taken altogether, then, the description of a cognitive system as proposed seems to satisfy, *prima facie*, the catalogue of constraints under which it is required to operate. Perception may be regarded, phenomenologically speaking, as passive, cognition active; the old intuitive Kantian division between receptivity and spontaneity in human thinking is, after a fashion, thereby preserved. Under normal conditions, new experience is incorporated smoothly into the existing corpus of thought in which active cognition takes place. When a conflict arises, second-order processes may be brought in to deal with the difficulty through the familiar application of justifications, explanations, and rules for decision, the flexibility of those decisions being parasitic on the rules and methods employed; these rules and methods themselves being open to revision just as much as the individual thoughts in the corpus. Belief remains a part of the human cognitive picture, thereby satisfying our intuition that some things may be believed rather than merely held probable or held as mere conjectures; yet it does not hold the central and highly problematic position that the philosophical tradition, whether tacitly or explicitly, has generally supposed it to hold. A significant proportion of the activity of cognition in the model may be carried on without ever adverting to the epistemic evaluation of individual thoughts or perceptions, which activity nonetheless may issue in rational action; this fits well with our introspectively apprehended phenomenal experience of our own cognition.

Lastly, although the mind in the model proposed may well possess metatheoretic theses concerning its own fallibility and the correctness of certain skeptical arguments, it does not require any such notions in order to operate according to specifications, and hence may qualify as descriptive of the cognitive processes of an unreflective, uncritical or juvenile mind.

Despite this last-minute attempt to tie the loose ends together, there are no doubt many problems with the description proposed which have been overlooked; similarly for the problem which sparked the generation of the description. The two portions of the paper may be taken separately; the description may survive without the problem which led to its construction. Even if the problem as stated is revealed, in the last analysis, to have been a pseudo-problem after all, its articulation goes some distance towards helping articulate some of the issues we may find ourselves entangled with when we consider the structure of the concept of belief. If the proposed description intended to overcome the problem is found to be internally flawed, I think it might stand some further refinement before it should be abandoned.

As for overcoming the skeptical worries which continue to guide my inquiries – I see no particular hope of eliminating these, although the possibility is always there. Standing on Neurath's boat, when the sun is shining, the wind good, and the water smooth, it may be difficult to imagine that there is no sun, no wind, and no sea; difficult also, to countenance the notion that the ship and the open ocean were not made for each other. When the skies darken and the seas grow heavy, and we can hear the planks and stays begin to buckle and crack, we may be more inclined to wonder whether we are beset by some evil influence – either one determined to destroy us, or one which earlier

misled us into thinking that the initial construction of our ship was sound; or worse, that our ship is in the condition of the Flying Dutchman, closely surrounded by a storm somehow generated by the ship itself, while nothing persists beyond. I cannot imagine what arguments we might make to ourselves to dispel such worries, but as long as the ship holds together through our continuous efforts at repair, we can go on – or go down – trying.

BIBLIOGRAPHY

- Carnap, R., *Meaning and Necessity*. University of Chicago Press, 1956.
- Cherniak, Christopher, *Minimal Rationality*. Cambridge, Massachusetts: The MIT Press, 1992.
- Chomsky, Noam, "Noam Chomsky." *The Blackwell Companion to the Philosophy of Mind*. Samuel Guttenplan (ed.), Blackwell, 1995.
- Churchland, P.M., "Replies." *Philosophy and Phenomenological Research* Vol. LVIII No. 4, December 1998.
- Dancy, Jonathan, *An Introduction to Contemporary Epistemology*. Blackwell, 1985.
- Davidson, Donald, "Mental Events." *Essays on Actions & Events*. Oxford (The Clarendon Press), 1980.
- Dawkins, Richard, *The Selfish Gene*. Oxford University Press, 1989.
- Dennett, Daniel C., "Self Portrait." *Blackwell Companion to the Philosophy of Mind*. Samuel Guttenplan (ed.), Blackwell, 1995.
- , "Intentional Systems." *Mind Design*. John Haugeland (ed.), The MIT Press, 1988.
- , *Consciousness Explained*. Little, Brown, 1991.
- , and Kinsbourne, M, "Response to Glicksohn and Salter." *Behavioral and Brain Sciences*, vol. 18, no. 4, 1995, pp. 810-11.
- Dummett, Michael, "Truth (1959)." *Truth and Other Enigmas*. Gerald Duckworth & Co. Ltd., 1992.

- Edmonds, David and Eidinow, John, *Wittgenstein's Poker*. Faber and Faber, 2001.
- Evans-Pritchard, E.E., *Witchcraft Oracles and Magic Among the Azande*. Oxford (The Clarendon Press), 1937.
- Fodor, Jerry, "Banish DisContent." *Language, Mind and Logic*. Jeremy Butterfield (ed.). Cambridge University Press, 1986.
- , *A Theory of Content and Other Essays*. Cambridge, Massachusetts: The MIT Press, 1992.
- , *In Critical Condition*, The MIT Press, 2000.
- Free Online Dictionary of Computing (FOLDOC)*, <http://foldoc.doc.ic.ac.uk/foldoc/>.
- Gärdenfors, Peter, "Belief Revision: An Introduction." *Belief Revision*, P. Gärdenfors (ed.), Cambridge University Press, 1992.
- Hartshorne, C., and Weiss, P. (eds.), *Collected Papers of Charles Saunders Peirce*, Volume 1, 1931-1935.
- Hookway, Christopher, *Quine*. Polity Press, 1988.
- , *Peirce*. Routledge & Kegan Paul, 1985.
- Horton, Robin, "African Thought and Western Science." *Rationality*, Bryan Wilson (ed.), Blackwell 1970.
- Hume, David, *A Treatise of Human Nature*. P. H. Nidditch (ed.), Oxford University Press, 1978.
- Kipling, Rudyard, "How the Camel Got His Hump." *Just So Stories*. Doubleday, 1912.
- Lacey, A. R., *A Dictionary of Philosophy*. Routledge & Kegan Paul, 1976.
- Lewontin, R. C. 1970., "The Units of Selection." *Annual Review of Ecology and Systematics* 1:1-18.

- Malthus, Thomas Robert, *An Essay on the Principle of Population*. W. W. Norton & Company, 1976.
- Mamassian, Pascal, Landy, Michael, and Maloney, Laurence T., "Bayesian Modelling of Visual Perception." *Probabilistic Models of the Brain: Perception and Neural Function*. R. P. N. Rao, B. A. Olshausen & M. S. Lewicki (Eds.), Cambridge, MA: MIT Press, 2002.
- Marr, David, *Vision*. W. H. Freeman and Company, 1982.
- McDowell, John, *Mind and World*. Harvard University Press, 1996.
- Norton, David F., "An Introduction to Hume's Thought." *The Cambridge Companion to Hume*. Cambridge University Press, 1993.
- O'Hear, Anthony, "Fallibilism." *A Companion to Epistemology*. Jonathan Dancy and Ernest Sosa, eds., Blackwell, 1992.
- Popper, Karl, "Evolution and the Tree of Knowledge." *Objective Knowledge An Evolutionary Approach*. Oxford University Press, 1979.
- Quine, W.V., and Ullian, J. S., *The Web of Belief* (1st edition). Random House, 1970.
- Quine, W.V.O., "Facts of the Matter." *Essays on the Philosophy of WV Quine*, Shahan and Swoyer (eds.), University of Oklahoma Press, 1979.
- Quinton, Lord A., "Knowledge and Belief." *Encyclopedia of Philosophy*. Paul Edwards (ed.), Collier-MacMillan, 1968.
- Rescher, Nicolas, "Fallibilism." *The Oxford Companion to Philosophy*. Ted Honderich (ed.), Oxford University Press, 1995.
- Ryle, Gilbert, *The Concept of Mind*. Hutchinson & Co., 1959.
- Sellars, Wilfrid, *Empiricism and the Philosophy of Mind*. Harvard University Press, 1997.

Stephan, Volker, and Gross, Horst-Michael, "*Neural Anticipative Architecture For Expectation Driven Perception*", presented at the 2001 Systems, Man, and Cybernetics Conference.

Stolc, Viktor, et. al., "A Gene Expression Map for the Euchromatic Genome of *Drosophila melanogaster*." *Science Magazine*, Vol. 306, No. 5696, 2004.

Winch, Peter, "Understanding a Primitive Society." *Rationality*. Bryan Wilson (ed.), Blackwell, 1970.