PAPERS ON PRAGMATISM

by

THOMAS MARK EDEN DONALDSON

A dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Philosophy
written under the direction of
ERNEST LEPORE
and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October 2012

ABSTRACT OF THE DISSERTATION

Papers on Pragmatism

by THOMAS MARK EDEN DONALDSON

Dissertation Director:

ERNEST LEPORE

Chapter One: James is often accused of claiming that a belief is true just in case it is useful. The objections to this view are obvious. I offer a more sophisticated interpretation of James's theory of truth, and defend it from the standard objections.

Chapter Two: I discuss Steve Stich's notorious claim that 'once we have a clear view of the matter, most of us will not find any value, either intrinsic or instrumental, in having true beliefs.' I argue that Stich reaches this conclusion only because he makes some false assumptions about content-determination. I show that using an interpretationist account of content-determination we can explain the value of true beliefs.

Chapters Three and Four: 'Definitionalism' is my name for the thesis that the theorems of mathematics are *a priori* because they are entailed by definitions. I discuss two objections to this view: the 'Kantian' objection and the Quinean objection. According to the 'Kantian' objection, definitionalism must be false because existential generalisations can't be true by definition. According to the Quinean objection, the definitionalist's distinction between definitions and other statements is illicit, because it is not discernible in mathematical practice. I argue that the Kantian objection is misguided, but the Quinean objection is correct.

Chapter Five: Elitists draw a distinction between two sorts of word: the elite and the plebeian. They claim that only by using the elite words can we describe the world as it is 'anyway'; using the plebeian words, we can at best describe the world as it seems to us, with our own particular physiology, tastes and history. For the sake of argument I grant these claims, but then contend that we have no way of identifying the elite expressions.

Acknowledgements

I'll begin with some personal acknowledgements. I would like to thank Mercedes Diaz for answering all of my foolish questions over the years; Bill Child for his help when I was an undergraduate; Brian Weatherson and Ernest Lepore for their advice on getting through grad school; Steve Stich and Holly Smith for teaching teaching; and all of the grad students at Rutgers for their friendship. Finally, I would like to thank my family for their support in my long absence.

Now I'd like to acknowledge people who have helped me with the philosophical work that follows. Andy Egan introduced me to contemporary metaethics; Cian Dorr made me think about the implications of the 'truth based view' (see Chapter 3) outside the philosophy of mathematics and alerted me to some of the problems with the notion of 'epistemological virtue' (see Chapter 5); Barry Loewer and Steve Stich helped me with Chapter 2; Frankie Egan showed me a whole new way of thinking about content-determination; Thony Gillies and the students in my dissertation seminar (Richard Dub, Rodrigo Borges, Justin Sharber, Ben Levinstein, and Thomas Blanchard) showed me some of the failings of the first draft of Chapter 5; Jonathan Schaffer's comments on sections 7 and 8 of Chapter 5 led to substantial improvements in the paper, and he showed me that not all metaphysicians are elitists; Jason Stanley showed me that it is not only people on the lunatic fringe who doubt elitism, or think that practical factors help to determine which beliefs constitute knowledge; Martha Bolton taught me about the British empiricists who preceded James; Bill Child's tutorials on Davidson got me interested in interpretationist theories of content-determination; Ted Sider told me about 'only mildly sceptical' forms of elitism; Kit Fine alerted me to an important lacuna in the appendix to Chapter 3; Alex Paseau introduced me to the philosophy of maths; Ian Rumfit and Crispin Wright got me interested in neofregeanism; Tobias Wilsch made comments on Chapters 1 and 5 which led to substantial improvements; the members of the Rutgers/Princeton

philosophy of maths reading group—Ben Levinstein, Jack Woods and Noel Swanson—taught me about indeterminacy in mathematics; Marco Dees, Meghan Sullivan and Jenn Wang have helped to understand how metaphysicians think; grad students who came to my Rutgers grad talks have helped me a lot, and in particular Nick Beckstead and Michael Johnson have asked me some important and difficult questions; Ben Levinstein's comments on Chapter 4 led to substantial improvements in its presentation; Bob Beddor taught me how to spot the 'conditional fallacy', which helped with Chapter 4; Carlotta Pavese has helped me clarify my thoughts on many topics over the years. Jenn Wang has read and helped me with everything that follows.

Finally, I'd like to thank Brian Weatherson and Ernie Lepore, for everything.

Contents

Introduction

Pragmatism and 'destructive naturalism'

By now the pattern should be familiar. Philosophers begin to worry that some of our ordinary, commonsensical beliefs don't cohere with what they take to be the scientific worldview. They argue about it for a while, and then some naturalist recommends eliminativism as a solution.

'Eliminativism' is not an easy term to define, but examples are not hard to find:

- Paul Churchland has drawn a comparison between the words of folk propositional attitude psychology ('believes', 'hopes', 'intends' and so on), and words from failed scientific theories, like 'phlogiston' and 'impetus'. On his view, we should reject 'Galileo believed that the world moves' and 'Hobbes hoped to square the circle' just as we reject 'Planets have circular impetus' and 'Phlogisticated air doesn't support life'. To put it colloquially, there is just no such thing as believing or hoping, just as there is no such thing as phlogiston or impetus.[1]

- Mackie said similar things about ethical terms like 'morally good', 'morally permissible', and 'morally virtuous'. According to Mackie, ordinary ethical assertions like 'it is wrong to enjoy the suffering of others' 'include ... an as-

[1] Churchland (1981).

sumption of objective values', but because 'there are no objective values,' such claims are 'all false'.[2]

- Since at least the eighteenth century, the claim that we lack free will has been sporadically defended by naturalistic philosophers.[3]

- In his book *Science Without Numbers*,[4] Hartry Field argued that there are no mathematical objects: no numbers, no sets, no functions etc.. For Field, the majority of mathematical 'theorems' are false. Contrary to what you've been told, there are not infinitely many prime numbers; and two doesn't have a square root.

- Maund defends the view that, while most physical objects appear to be coloured, this appearance is not veridical. Bananas look yellowy-green but they aren't; also walnuts aren't brown, and elephants aren't grey.[5]

It's easy to see why people sometimes think of naturalism as a threatening movement. These claims, taken together, are both *wildly* counterintuitive and threatening to our sense of what's important. The challenge is to rebut 'destructive naturalism', as I will call it, without resorting to superstition.

A pragmatist response to this challenge has two parts. First, the pragmatist defends our use of the supposedly problematic concepts—modal, aesthetic, mathematical, or whatever—by appeal to their utility. Second, when asked whether we should be 'realists' about out these concepts, the pragmatist attempts somehow to undermine or trivialize the question. The pragmatist hopes thereby to defend the

---

[2]Mackie (1977). Joyce defends a similar view in Joyce (2001).

[3]I won't attempt a thorough list of citations! Two recent examples are Perebroom (2006) and Wegner (2002).

[4]Field (1980).

[5]Maund (2006).

problematic beliefs from the attacks of the destructive naturalists; at the same time, so long as the explanation of the utility of the beliefs in question is suitably naturalistic, no superstition need be involved.

The papers in my dissertation are on different topics, but each is a part of an extended investigation of this pragmatist idea. In this introduction, I summarise the papers and explain how they fit together.

James

'The present dilemma'

In the first lecture of *Pragmatism*,[6] James distinguished two sorts of philosopher: the 'tough-minded' and the 'tender-minded'. Roughly, 'tough-minded' means 'naturalistic' and 'tender-minded' means 'religious', but a slightly more detailed description of the two terms will be useful.

When James talked about the 'tough-minded', he had in mind philosophers who are empiricist in their epistemology, and whose metaphysical views are motivated by scientific considerations. Quine would be a good recent example, though of course he worked after James. Nowadays we expect tough-minded philosophers to be physicalists, and indeed James said that tough-minded philosophers are often 'materialistic'. However, he put many phenomenalists in the same category.[7]

James found the metaphysical views of the tough-minded too austere; he thought that the tough-minded omit much that is of importance to us, or offer unsatisfying naturalistic reductions.

---

[6] James (1907).

[7] The idea that phenomenalism is a tough-minded position seems very odd to us today—but among James's predecessors there are some phenomenalists who do seem to deserve the title 'tough-minded'. Mach (Mach (1984)) and Clifford (Clifford (1878)) are examples.

As he put it:

> [T]he view is materialistic and depressing. Ideals appear as inert by-products of physiology, what is higher is explained by what is lower and treated forever as a case of 'nothing but'—nothing but something else of a quite inferior sort. You get, in short, a materialistic universe, in which only the tough minded find themselves congenially at home.[8]

When James talked about 'tender-minded' philosophers, he had in mind primarily rationalists who endorsed a form of monism, according to which 'the world is no collection, but one great all-inclusive fact outside of which is nothing'.[9] The typical tender-minded philosopher claimed that this 'all-inclusive fact' is conscious, infinite, morally perfect and unchanging. Josiah Royce is a clear example;[10] James put the British idealists in the same category.

James liked their attempt to describe a universe in which we can feel 'at home', but nevertheless he was rather dismissive of the views of the tender-minded. He claimed that the tender-minded philosophers had a 'feeble grasp of reality', and so they constructed 'pure but unreal theories' that are 'out of touch with concrete facts, and joys and sorrows'. More seriously, he claimed that the tender-minded views are discredited by a version of the problem of evil: in claiming that the 'absolute thought' is morally perfect, James thought, the tender-minded fail to take seriously the reality of suffering.

Some of the time, James was attracted to an extremely tough-minded, phenomenalistic position.[11] He occasionally advocated a view which I call 'hard phenomenalism', according to which the only objects that exist are sensations, and the only

---

[8]From Lecture I of James (1907). James seems took the term 'at home' from Hegel, who said that 'the aim of knowledge is to divest the objective world of its strangeness, and to make us feel more at home in it'. James agreed. See the first lecture of James (1909).

[9]From 'The Types of Philosophic Thinking', in James (1909).

[10]Royce (1885).

[11]See for example James (1904a) or James (1912b).

properties that these things have are phenomenal properties. But in *Pragmatism*, James rejected this position. His goal was to construct a theory which would make him feel 'at home'; at the same time he wanted to avoid the airy fantasies of the tender-minded by maintaining a basically phenomenalist orientation:

> You want a system that will combine both things, the scientific loyalty to facts and willingness to take account of them, the spirit of adaption and accommodation, in short, but also the old confidence in human values and the resultant spontaneity, whether of the religious or of the romantic type. And this is then your dilemma: you find the two parts of your quaesitum hopelessly separated. You find empiricism with in-humanism and irreligion; or else you find rationalistic philosophy that indeed may call itself religious, but that keeps out of all definite touch with concrete facts and joys and sorrows.[12]

James's solution

When describing James's attempted solution, I like to start with his developmental psychology. James thought that new-born babies have only phenomenal concepts. In consequence, their perceptions seem totally disordered: a 'blooming, buzzing confusion'.[13] As children grow up, they acquire non-phenomenal concepts and beliefs which they use to systematise their perceptual input. For example, children learn to group similar sensations together to form persisting objects; they then learn to classify these objects in useful ways. They learn to think of these objects as distributed through space, which has metrical properties which remain constant as time passes. They learn to think of time as measurable on a scale, so that rates of change can be numerically compared. And so on. The result is that by the time we are adults our experience no longer seems to be a mere tangle of miscellaneous sensations. On the contrary, it is richly patterned.

---

[12]James (1907), Lecture I.

[13]This famous phrase comes from Chapter XIII of the first volume of James (1890).

So James maintained the idea that we live in a 'world of pure experience'. At the same time, he hoped to justify our non-phenomenal beliefs by saying that they serve to to systematise the underlying phenomenal truths. Our goal as believers is to find systematisations which help us achieve our goals.

This is how James wanted to solve the tough/tender problem. On the one hand, he maintained a 'tough', phenomenalist orientation, and by focusing on practical utility James thought that he could avoid the airy fantasies of the tender-minded. At the same time, James hoped to defend our non-phenomenal beliefs (mistakes aside, of course) by appeal to their practical value in systematising the underlying phenomenal facts. In particular, James thought, our religious moral and aesthetic beliefs play a pragmatically valuable role in our system of belief, and are in consequence justified.

James's views on truth

It's tempting to object that James has so far not managed to present an alternative to the hard phenomenalist position. After all, even the most tough-minded philosophers should be ready to agree that our non-phenomenal beliefs are *useful*. Surely, if James's position is to be a real alternative to the tough-minded views, he must argue that our non-phenomenal beliefs are not only useful, but in most cases true.

This is where James's theory of truth comes in.

In some passages, James advocated the crude view that a belief is true just in case it is useful—a claim which has attracted well-deserved ridicule ever since.[14] In the first paper of this dissertation, I argue that in these passages James was simplifying his position. James's theory truth, in more detail, is that a belief is true just in case it is an element of the best systematisation of the underlying phenomenal truths. I

[14]For example, see the second lecture of James (1907), where James writes, 'I am well aware of how odd it must be for some of you to hear me say that a belief is true so long as to believe it is profitable for our lives'.

argue that this version of James's theory is not refuted by the familiar point that some truths are useless, or by the point that sometimes it is useful to believe things which are false.

Clearly, few philosophers today will be attracted to James's theory of truth, because few philosophers today have any sympathy of James's phenomenalism. However, if you replace James's phenomenalism with physicalism, the resulting position has some contemporary appeal. I conclude my paper by giving some reasons for rejecting this updated version of James's view.

The value of truth

James, as I've said, sometimes identified the truth of a belief with its utility. In *The Fragmentation of Reason*,[15] Stich adopts a superficially very different position. 'Once we have a clear view of the matter,' he argued, 'most of us will not find any value, either intrinsic or instrumental, in having true beliefs'.[16]

However, despite their differences, Stich and James agree on some important points. Stich describes his position as a form of pragmatism, and the label is well chosen. This is what Stich has to say about the goal of inquiry:

> But if truth is not to be the standard in epistemology, what is? The answer that I favor is one that plays a central role in the pragmatist tradition. For pragmatists, there are no special cognitive or epistemological values. There are just values. Reasoning, inquiry and cognition are viewed as tools that we use in an effort to achieve what we value. And like any other tools, they are to be assessed by determining how good a job they do at achieving what we value. So on the pragmatist view, the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that he or she finds intrinsically valuable.[17]

---

[15]Stich (1990), chapter 5.

[16]Stich (1990), pg. 108.

[17]Stich (1993).

James would no doubt agree that 'the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that he or she finds intrinsically valuable',[18] though he would add that the beliefs selected by these 'good cognitive strategies' will typically be *true*, precisely because they help one achieve what one values.

Before getting to Stich's argument, I should discuss an ambiguity in the word 'belief'. Consider this sentence:

> My love gets bigger by the day.

This sentence is ambiguous. On one reading, it is a declaration of ever increasing love; on the other, it is a declaration of love for one who is ever increasing. This shows that the word 'love' is ambiguous. It can be used to refer to the person who is loved, or it can be used to refer to the state of the lover. The word 'belief' is similarly ambiguous: one can use it to refer to the *proposition believed*, or to the *state of the believer*.[19]

Following philosophical orthodoxy, Stich supposes that there is a function which maps belief-states to the corresponding propositions,[20] and that a belief-state is true just in case this function maps it to a true proposition.

I call this function 'the standard interpretation function'. Of course, there are

---

[18]There's a 'choice point' here which I am ignoring to avoid excess complexity in the discussion. Contrast:

> Stich: the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that he or she finds intrinsically valuable.
>
> Alternative: the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that are, objectively, intrinsically valuable.

I don't know which of these accounts James would have preferred.

[19]I am confident that this ambiguity exists in philosophical English. I'm not so sure about whether non-philosophical English has the same ambiguity.

[20]To account for vagueness, it is perhaps better to suppose that there are many such functions, corresponding to different ways of precisifying a person's beliefs. I'm ignoring this complication—it doesn't affect the argument in any important way at this stage.

many other functions which map any given person's belief states to propositions. I'll call these 'alternative interpretation functions'. Let's call the standard interpretation function 'I_0', and let I*, I**, I*** and so on be alternative interpretation functions. Then just as a belief-state is called 'true' if I_0 maps it to a true proposition, we can say that a belief-state is TRUE* if I* maps it to a true proposition, or TRUE** if I** maps it to a true proposition, and so on. Stich's argument begins with the question: do we have any reason to think that true belief-states are of any more value than TRUE* ones, or TRUE** ones (etc.)? What makes the standard interpretation function special, from a normative point of view?

Stich can find *no* reason for thinking that the standard interpretation function is special, or for thinking that seeking truth is more advisable that seeking TRUTH*, or TRUTH**.

In 'Does the Truth have Cash Value?', I discuss the existing responses to Stich's argument in the literature, and argue that they are inadequate. Then I give my own response. Briefly, I argue that the reason that Stich is unable to find anything special about the standard interpretation function is that his search is hampered by a mistaken theory of content-determination—that is, a mistaken theory about what qualifications an interpretation function must have in order to deserve the title 'the standard'. I recommend instead an 'interpretationist' account of content-determination, rather like that suggested by David Lewis in his paper 'Radical Interpretation',[21] and show that using this theory of content determination one can explain the specialness of the standard interpretation function.

---

[21]Lewis (1974).

Mathematics

Mathematics as a case study for pragmatists

Here are two apparent features of mathematics which naturalists sometimes find problematic. First, mathematics seems to concern non-physical things. Second, mathematics seems to be an *a priori* subject: mathematicians don't tend to provide empirical evidence for their claims, and they don't seem to think that pure mathematical claims can be refuted by experiential data. The first feature of mathematics is problematic for the many naturalists who incline towards physicalism. The second feature is problematic for the many naturalists who incline towards empiricism.

It is not surprising, then, that some naturalists have responded by rejecting mathematics, in whole or in part. Quine is a case in point. A convinced empiricist, Quine thought that we are justified in accepting only those mathematical statements which are entailed by well-established scientific theories. So Quine was inclined to reject those parts of pure mathematics which are not used in the sciences—higher set-theory for example.[22]

In the past few decades, mathematicians who study 'reverse mathematics' have shown that remarkably weak mathematical systems suffice for much of the mathematics that is used in the sciences.[23] It is plausible that all of the mathematics needed in the sciences can be reconstructed within second-order arithmetic. If this is right, a consistent Quinean would have to have to reject much of ordinary set-theory.

Hartry Field's views are still more extreme than Quine's.[24] Field recommended a view on which there are no mathematical objects: no sets, no numbers, no functions,

---

[22]In Quine (1992), Quine recommends the Axiom of Constructibility, saying that '[i]t inactivates the gratuitous flights of higher set theory'.

[23]See Simpson (2010) for a summary.

[24]Field (1980).

and so on.

Field and Quine's views about mathematics exemplify the sort of destructive naturalism that I described in the opening section of this introduction. It provides a useful case study for pragmatists.

Carnap's pragmatist views about mathematics

I am not saying that Carnap was a pragmatist: it is useful to separate the pragmatists from the logical empiricists in our philosophical taxonomy. However, there are undoubted similarities between the views of philosophers in the two schools, and Carnap's views about the ontology of mathematics, as he described them in 'Empiricism, Semantics and Ontology',[25] exactly fit the pattern I described in the first section of this introduction.

Carnap began the paper by saying that '[e]mpiricists are in general rather suspicious with respect to any kind of abstract entities' and explaining that his goal in the paper is to show that 'using [a language referring to abstract entities] ... is perfectly compatible with empiricism and strictly scientific thinking'. He discusses both mathematics and semantics, but I'll stick to the mathematical case.

Carnap began his presentation of his views by outlining a simple account of the semantics of mathematical words. Carnap explained that our use of mathematical words is rule-governed: there are syntactic rules, and there are semantic rules. The semantic rules, he said, are 'deductive' —i.e. they specify which deductions involving mathematical terms are valid. For example, the semantic rules for the vocabulary of number theory might allow one to assume each of the Peano axioms at any stage in a proof.

Carnap then discussed 'realism' about number theory, though he didn't use the

[25]Carnap (1950).

word. He distinguished two questions, which he called the 'external question' and the 'internal question'. The external question is 'Should we use the number-theoretic vocabulary?'; the internal question is 'Given these semantical rules, is it true that there are numbers?'. He claimed that the first question is to be settled on pragmatic grounds: we should use number-theoretic vocabulary if doing so is sufficiently useful, when compared to the relevant alternatives. And Carnap indicated that he did think that number-theoretic vocabulary has sufficient utility that we should continue to use it. Turning to the second question, Carnap claimed that the answer is trivially 'Yes'. 'There are numbers' is a more or less immediate consequence of the semantical rules governing the number-theoretic terms. For Carnap, it is just analytic, or true by definition, that there are numbers.

'Definitionalist' is my term for philosophers who, like Carnap, think that mathematical theorems are knowable because they are entailed by definitions.[26] There are two main objections to this 'definitionalist' position in the literature: one Kantian, one Quinean. According to the Kantian objection,[27] definitions can't entail existential generalisations, and so it cannot be true in general that mathematical theorems are entailed by definitions. According to the Quinean objection, the distinction between definitions and other statements is illicit, because no such distinction is drawn in mathematical practice.[28] As I will explain in the next two sections, I think that

---

[26]James accepted definitionalism, or something very like it. In the sixth lecture of James (1907), he wrote:

> But matters of fact are not our only stock in trade. Relations among purely mental ideas form another sphere where true and false beliefs obtain, and here the beliefs are absolute, or unconditional. When they are true they bear the name either of definitions or principles. It is either a principle or a definition that 1 and 1 make 2, that 2 and 1 make 3, and so on; that white differs less from gray than it does from black; that when the cause begins to act the effect also commences.

[27]See Sider (2007) for a nice presentation of the objection, with some discussion.

[28]See Quine (1953) for Quine's views on definitions.

the Kantian objection is weak, but the Quinean objection is decisive.

The Kantian objection to definitionalism

There is a remarkable divergence of intuitions on the question of whether existential generalisations can be analytic. On the one hand, some people find the view that the theorems of mathematics are analytic highly plausible. Other people think it preposterous to say that an existential generalization is analytic. In my third paper, I attempt to explain this divergence of intuition, and defend the definitionalists against the Kantian objection.

Let's look at the Kantian objection in a little bit more detail. First, notice that sometimes definitions fail. For example:

> a is the largest prime number.
>
> John is the man who assassinated Oscar Wilde.
>
> Phlogiston is the substance that is released during combustion.

The Kantian explains the failure of these definitions in the following very natural way. The point of the definitions, the Kantian says, is to assign referents to the newly introduced terms (viz. 'a', 'John' and 'phlogiston'). Now as it happens there don't exist suitable referents for these terms, and so the definitions fail. You can't stipulatively assign the name 'John' to someone who assassinated Oscar Wilde, because there is no such person.

Now let's consider mathematics. Suppose that one gives stipulative definitions for the basic vocabulary of number theory, including the numerals '0', '1', '2', '3' and so on. According to the Kantian, one's definitions will succeed only if there exist suitable referents for all of these numerals.

Given that all of these are theorems of number theory:

$0 \neq 1$			
$0 \neq 2$	$1 \neq 2$		
$0 \neq 3$	$1 \neq 3$	$2 \neq 3$	
$0 \neq 4$	$1 \neq 4$	$2 \neq 4$	$3 \neq 4$
...	...	...	...
...	...	...	...

there will exist suitable referents for all of the numerals only if there exist infinitely many things. If the universe is finite, the Kantian claims, one's number theoretic definitions will fail—like the definitions of 'a', 'John' and 'phlogiston' considered above.

The Kantian argues that, *pace* the definitionalist, one cannot establish the existence of the natural numbers just by deducing this conclusion from some stipulative definitions. Even if one's definitions of the vocabulary of number theory entail that 0, 1, 2, 3 and so on exist, one cannot know on this basis that these numbers exist without some independent assurance that one's definitions have succeeded; in particular, one would need some reason for thinking that the universe is not finite.

This objection to definitionalism makes use of a certain theory about definitions, which I call the 'reference based theory'. According to the reference based theory, the function of a set of definitions is to assign referents to the newly introduced terms, and the definition will succeed just in case there exist suitable referents.

Definitionalists themselves tend not to think of definitions in this way. According to an alternative theory—what I call the 'truth based theory'—the function of a definition is to assign truth-conditions to *whole sentences* containing the newly defined expression or expressions.

In the third paper I explain the 'truth based' and 'reference based' theories in more detail. I argue that the reference based theory is indeed inconsistent with definitionalism. Finally, I develop in detail a version of the truth based theory, which I use to defend definitionalism against the Kantian objection.

The Quinean objection

According to the barroom version of the Quinean objection, definitionalism fails because there is no way to draw the distinction between definitions and other truths. This version of the Quinean objection can be dismissed quickly. There are in fact lots of ways of drawing this distinction. For example, we could draw the distinction by saying that the definitions are all and only those sentences labeled 'definition' in a standard mathematics textbook.

A better way of putting the Quinean objection is that, while there are several ways of drawing a distinction between definitions and other truths, none of these distinctions can play the theoretical role that the definitionalist intends.[29] To return to my example, suppose we use the word 'definition' so that the definitions are all and only those sentences labeled 'definition' in a standard mathematics textbook. If the definitionalist uses this interpretation of 'definition', she will have to accept the surely undesirable conclusion that some contradictions are true, because the definitions that appear in standard mathematics textbooks are not consistent.[30]

In my paper on the Quinean objection to definitionalism, I go through various different ways of understanding the term 'definition', arguing in each case that definitionalism fails.

---

[29]See Williamson (2008) for a similar complaint.

[30]I suppose that some philosophers will be willing to accept that some contradictions are analytic, as part of a response to the semantic paradoxes. But surely even a dialetheist will not want to accept that natural numbers both are and are not integers, as this version of definitionalism implies. See the fourth paper for more details.

Elitism

'Elitists' are those who distinguish two sorts of word—the elite and the plebeian.[31] For example, it is said that words from logic, maths and fundamental physics are elite, while aesthetic terms, and secondary quality terms are plebeian.

The idea is that the elite words are somehow better fitted to the structure of the world, so that one of our goals when constructing theories should be to use elite terms. When using elite words, so the elitists claim, we can describe the world 'as it is anyway' rather than merely 'as it appears to us'.

As I say, many elitists think that some logical words are elite; in particular, it is often said that there is an elite existential quantifier. This idea can get you out of some tight spots when doing ontology. Here's an easy example. Consider the sentence:

> 'There is a hole through every ring doughnut.'

A metaphysician might find herself conflicted when considering this sentence. On the one hand, she might have some theoretical reason for denying that holes exist. On the other, the sentence seems obviously true.

The elitists have a solution. An elitist can say that the English sentence 'there is a hole through every ring doughnut' is true, but only because the English quantifier 'there is' is plebeian. When talking about metaphysics, she will say, we should avoid this clumsy quantifier, and use an elite quantifier—'there *really* exists', say—instead. She can then claim that this elite quantifier does not range over holes. So the metaphysician can have it both ways. She can say, with the folk, that there is a hole through every ring dougnut; at the same time, she can insist that there don't *really* exist holes.

[31] Sider (2011) is the clearest example.

Let's return to number theory, my case study. Suppose the pragmatist, or whoever, manages to show that the theorems of number theory are genuine statements (rather than mere 'moves in a formalism') and that they true. My view is that this suffices to establish realism about number theory. The elitist, however, is likely to say that realism has not yet been established: there remains the question of whether numbers *really* exist.

More generally, I said back in the opening section of this introduction that pragmatists characteristically try to undermine or trivialize questions of realism. So you might expect pragmatists to be suspicious of people who distinguish the question of what exists from the question of what really exists. And indeed, recent pragmatists, including both Putnam and Rorty, have rejected elitism.[32,33]

In my last paper, 'Against Elitism', I discuss the epistemology of elitism: granting for the sake of argument the distinction between elite and plebeian words, how can we identify the elite ones? I know of only one serious answer to this question, due to Sider. He explains his proposal in this passage:

> Quine's advice for forming ontological beliefs is familiar: believe the ontology of your best theory. Theories are good insofar as they are simple, explanatorily powerful, integrate with other good theories, and so on. We should believe generally what good theories say; so if a good theory makes an ontological claim, we should believe it. The ontological claim took part in a theoretical success, and therefore inherits a borrowed luster; it merits our belief. This all is familiar; but a believer in [eliteness] can say more. A good theory isn't merely likely to be true. Its ideology is also likely to [be elite]. For the conceptual decisions made in adopting that theory—and not just the theory's ontology—were vindicated; those conceptual decisions also took part in a theoretical success, and also inherit a borrowed luster.
>
> . . .

[32]See 'Against Elitism' for citations.

[33]What about the pragmatists of the first wave? I am inclined to think that both James and Peirce were elitists. It's plausible to attribute to James the claim that phenomenal vocabulary is elite. Peirce called himself a 'scholastic realist', because he believed in what we now called 'sparse properties'—a view closely related to elitism.

> The Quinean thought about ontology is sometimes put in terms of indispensability: believe in the entities that are indispensable in your best theory. The analogous thought about ideology may be similarly put: regard as joint-carving the ideology that is indispensable in your best theory. This is fine provided "indispensable" is properly understood, as meaning: "cannot be jettisoned without sacrificing theoretical virtue."[34]

In my paper, I argue by giving examples that with sufficient ingenuity we can dispense with any word, and so by Sider's criterion we are never in a position to know that a term is elite.

[34]Sider (2011) Section 2.3.

Chapter 1

A Reassessment of James's Theory of Truth

Most philosophers today don't take William James's theory of truth very seriously. I think this is a shame. I don't accept James's theory, but I do think that there's more to be said in its favour than people usually think.[1] It deserves a better reputation.

In section one I'll briefly summarise the familiar case against James's theory of truth. In section five I'll present my interpretation of his theory, having prepared by explaining some of James's other philosophical views in sections two, three and four. Then I'll evaluate the theory. I'll focus on *Pragmatism*, but I'll need to draw upon some of James's other works too. His earlier views won't be considered.

---

[1]Part of the problem is that the lectures are light on detail, because James was writing for a popular audience. 'I have no right to assume that many of you are students of the cosmos in the classroom sense', James wrote in the first lecture of James (1907). From now on, all quotations from James are from this work unless I state otherwise.

1.1 The case against James's pragmatist theory of truth

James sometimes said that a belief is true just when it is 'profitable' to have that belief.[2] It is obvious that this can't be right. One or other of the following is true:

(A) There were evenly many grains of rice in my rice jar at noon yesterday.

(B) There were oddly many grains of rice in my rice jar at noon yesterday.

However, nobody would profit from believing either one of these things. In the other direction, sometimes it is useful to believe a falsehood; for example, it may benefit a shy man on a date to believe that he is handsome, even if he isn't.

At other times James sounded more Peircean, and seemed to endorse the view that a belief is true if and only if it would be believed at an imaginary time in the future when inquiry has finished. This too cannot be right, because of what Peirce called 'buried secrets'. Either (A) or (B) is true, but since I ate some of the rice from my jar last night, nobody will ever know which. The information is lost forever. Were inquiry to finish, nobody would have an opinion on the topic.[3]

So it's obvious that neither of these theories of truth is correct. I think that it would be a mistake however to attribute either one of these silly theories to James: when one reads the surrounding passages carefully, one finds all manner of hedges and qualifications. The problem for the interpreter is that while James provided plenty of explanatory metaphors and rough statements, he never got around to giving a clear statement of his position. As a result, the reader can begin to feel that he has to

---

[2]For example, in his second lecture, James says 'I am well aware of how odd it must be for some of you to hear me say that a belief is 'true' so long as to believe it is profitable for our lives.'

[3]In Peirce (1878), Peirce wrote, 'The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth'. For a partial defence of Peirce's position, see Misak (2004).

choose between attributing some crude theory to James—a theory which is easily refuted—or concluding that James's 'theory' of truth is just a tangle of miscellaneous ideas.

Another major problem with James's claims about truth is that he associated them very closely with his idealism. He insisted that truth is a 'man-made product', and connected this with his claim that the world is 'plastic' rather than 'ready-made'. On a first reading of the text, James seems to have thought something like this. Assuming the pragmatist theory of truth, it is true that some dinosaurs were herbivores because it is useful to believe that some dinosaurs were herbivores. So our interests are part of what make it true that some dinosaurs were herbivores. Therefore, our interests affect the diets of dinosaurs. Needless to say, this line of thought is ridiculous. Many people react by saying, 'so much the worse for James's theory of truth'.

1.2 James's pure experience theory

In *The Religious Aspect of Philosophy*, Royce argued that only an absolute idealist can give an adequate account of the relation between an idea and its object (i.e. the thing that the idea is about).[4] Royce's argument made a big impression on James, who devoted much his philosophical effort over the following years to the development of an empiricist account of this relation.[5] In 'Does "Consciousness" Exist',[6] he offered a surprising solution to a special case of Royce's problem, by giving an account of the relation between an 'impression' (in Hume's sense) and its object. Suppose, for example, that I am looking at knife. According to James's empiricist theory of

[4]See chapter XI of Royce (1885) in particular.

[5]For historical discussion see Sprigge (1997).

[6]James (1904b).

perception, my impression is a sort of pointy-shaped colour patch in the visual field. The relation between this colour patch and the actual, physical knife, according to James, is just the mereological part-whole relation. For James, the knife is made out of sensations, one of which is my impression of it. So James rejected the dualist idea that the knife is made of physical stuff, while the knife-impression is made out of something else, some kind of mental stuff.

James didn't shy away from the Berkeleyan character of the theory; indeed, *Pragmatism* is full of admiring references to Berkeley, whom he regarded as a sort of pragmatist *avant la lettre*:

> Berkeley's treatment of matter is so well known as to require hardly more than a mention. So far from denying the external world which we know, Berkeley corroborated it. It was the scholastic notion of a material substance unapproachable by us, behind the external world, deeper and more real than it, and needed to support it, which Berkeley maintained to be one of the most effective reducers of the external world to unreality.[7]

James had a couple of reasons for accepting this theory about perception. First, it got him part-way to his goal of providing an empiricist account of the relation between an idea and its object. Second, as the passage just quoted suggests, he seems to have found it mysterious and unnecessary to suppose that there's a 'real' knife, hidden behind the thing we directly perceive.

An obvious question at this point is: What about objects that have never been experienced, and never will be? For example, what about rocks on the dark side of the Moon? Are they made of experiences too? James refused to follow Berkeley in appealing to theism here; instead, he took a hint from Mach and Clifford[8] and claimed that such objects are made out of experiences that are never had by anyone:

---

[7]This is from the third lecture.

[8]See Mach (1984) and Clifford (1878). Banks (2003) is a useful history of these ideas.

experienCES without experienCERS. James called this the 'pure experience theory'.[9]

So James thought that sensations are mind-independent, at least in the sense that there are sensations which are not had by anyone. James liked to emphasize the mind-independence of sensations, saying that they are 'found, not manufactured'. Furthermore, James thought that some of the properties of sensations, and relations between them, are mind-independent.[10] On James's view, minds are merely compounds of sensations, so minds are sensation-dependent, but sensations are not mind-dependent. I'll say more about what 'mind-independence' amounts to in section eight.

This probably sounds ridiculous. It's worth pointing out, in James's defence, that the theory sounds more plausible when described using different terminology. If James's 'experiences' can exist without experiencers, it's best not to call them 'experiences' at all. They could be called 'basic objects' or something else instead. Then you can summarize James's theory by saying that the objects of direct perception are basic objects, which are also the constituents of physical things. Described like this, James's theory sounds less like Berkeleyan idealism and more like Reid's direct

---

[9] As he puts it in James (1904a):

> we at every moment can continue to believe in an existing beyond. It is only in special cases that our confident rush forward gets rebuked. The beyond must, of course, always in our philosophy be itself of an experiential nature. If not a future experience of our own or a present one of our neighbor, it must be a thing in itself in Dr. Prince's and Professor Strong's sense of the term ...

Strong's definition of 'thing in itself' in Strong (1903) is this:

> By 'things-in-themselves' I understand realities external to consciousness of which our perceptions are the symbols.

So James's claim was that there are things 'of an experiential nature' which are nevertheless 'external to consciousness'—i.e. there are un-had experiences.

[10] The textual evidence for this reading is indirect, but decisive. In James (1904a) James claimed that some relations between sensations are themselves sensations. Since, on James's view, sensations are mind-independent, it follows that some relations between sensations are mind-independent.

realism: in fact, James played up this similarity from time to time.[11]

However, even using the new terminology, James's theory looks very peculiar to philosophers today. To see why, think about James's views on hallucination. Contrast two cases. In the first case, I have a very realistic hallucination of a knife; in the second, I have a veridical perception of a knife. We can suppose that the hallucination is so realistic that the two cases are not distinguishable. What's the difference between the two cases? According to James, a real knife is a compounded sequence of sensations (or 'basic objects'); in the first case, the sensation that I have is not part of the right sort of sequence, and so it's not part of a real knife. In the second case, the sensation that I have is part of the right kind of sequence, and so the experience is counted as veridical. So for James, the sensation that I have when I hallucinate is intrinsically identical to part of a real, physical knife. 'Mental knives may be sharp,' James wrote, 'but they won't cut real wood.'[12] This is hard to believe.

1.3 Some misleading passages

There are several passages in *Pragmatism* in which James seems to take back his claim that sensations mind-independent. An unwary reader might conclude that we should regard James's claim that sensations are 'found, not manufactured' as a slip. In this section, I'll look at two such passages and explain why they should not be taken to be inconsistent with James's claims about the mind-independence of the phenomenal.

---

[11]For example, in James (1904a), James writes: 'it is impossible to subscribe to the idealism of the English school. Radical empiricism has, in fact, more affinities with natural realism than with the views of Berkeley or of Mill'.

[12]From James (1904b).

Here's the first:

> even in the field of sensation, our minds exert a certain arbitrary force. By our inclusions and omissions we trace the field's extent; by our emphasis we trace its foreground and its background; by our order we read it in this direction, then in that. We receive in short the block of marble; we carve the statue ourselves.

Properly understood, this passage does not contradict James's claim that sensations are 'found, not manufactured'. Rather, in this passage James was objecting to what he sometimes called 'atomism'.[13] James thought that sensory input comes to us, not in the form of a set of ready-distinguished sensations (as an atomist would claim), but in the form of a continuous stream. He thought that selective attention is needed to identify a particular sensation. One picks out a single sensation by ignoring the rest of the stream of sensory input. This is the claim that James is making in this passage; he is not claiming that sensations are mind-dependent. We can see that this is what James had in mind by his choice of metaphor. James had used the sculptor metaphor before: both in 'The Spatial Quale'[14] and in his *Principles of Psychology*;[15] these earlier works can be used to disambiguate the later one.

Here is another passage in which James appears to withdraw his claim that sensations are mind-independent:

> When we talk of reality 'independent' of human thinking, then, it seems a thing very hard to find. It reduces to the notion of what is just entering into experience and yet to be named, or else to some imagined aboriginal presence in experience, before any belief about the presence had arisen, before any human conception had been applied. It is what is absolutely dumb and evanescent, the merely ideal limit of our minds.

---

[13]See Klein (2007).

[14]James (1879b).

[15]See James (1890), chapter XI vol. 1.

Again, we need not read James as contradicting his claim that sensations are mind-independent; he has something else in mind here. When James said that 'reality independent of human thinking is absolutely dumb', he was simply making the straightforward point that our sensations don't describe themselves, and that when we describe them, we have to choose what to emphasise. His claim that sensations are 'the merely ideal limit of our minds' is a little harder to interpret, but it makes sense when compared with his earlier work. In the *Principles of Psychology*, James distinguished 'sensation' from 'perception'. The stream of sensations, as I said, is to be thought of as the unprocessed input to the mind. This input is then supplemented by our cognitive activity, and the result is perception:

> From the physiological point of view both sensations and perceptions differ from 'thoughts' (in the narrower sense of the word) in the fact that nerve-currents coming in from the periphery are involved in their production. In perception these nerve-currents arouse voluminous associative or reproductive processes in the cortex; but when sensation occurs alone, or with a minimum of perception, the accompanying reproductive processes are at a minimum too.[16]

Now James claimed that in adults, the arousal of 'associative or reproductive processes in the cortex' is unavoidable, so 'pure sensation' is more or less impossible. 'Pure sensations can only be realised in the earliest days of life,' James claimed; 'they are all but impossible to adults with memories and stores of association acquired.' He illustrated this claim with an example. We English-speakers find it hard to hear the sentence 'Paddle your own canoe' without letting our knowledge of the meanings of the words influence our perception; in consequence, we find it difficult to recognise the similarity of sound between 'Paddle your own canoe' and the French 'Pas de lieu Rhône que nous'.[17]

---

[16]From James (1890) vol. 2, chapter XVII.

[17]Again, this is from James (1890) vol. 2, chapter XVII.

I suggest that this is what James had in mind when he said that a 'reality independent of human thinking' is 'the merely ideal limit of our minds'. An adult can try to stifle the tendency for his sensations to provoke associations, and in doing so he may get close to the sort of pure sensation of which neonates are capable, but he may never wholly succeed.

1.4 James on the function of non-phenomenal concepts

So James advocated a form of phenomenalism. He thought that physical things are made out of sensations. At times, he went further, advocating what I'm going to call 'hard phenomenalism', the view that everything is phenomenal. On this view, all objects are sensations, the only properties they have are phenomenal properties,[18] and the only relations they bear to one another are phenomenal relations.[19] In *Pragmatism* however, James rejected hard phenomenalism, calling this the 'tough-

[18]I should say something about what 'phenomenal' means in this context. A good first pass at a definition would be: 'The phenomenal properties/relations are those that are given in experience—i.e. those that can be perceived without any processing'. But this isn't quite right, because James thought that attention is needed to separate any part of experience from the rest (see section 3). We should say instead that the phenomenal properties are those which can be perceived without any kind of processing apart from that involved in selective attention.

[19]Here are a couple of quotations to this effect. The first comes from James (1904a):

> I give the name of 'radical empiricism' to my Weltanschauung. ... To be radical, an empiricism must neither admit into its constructions any element that is not directly experienced, nor exclude from them any element that is directly experienced.

The second comes from James (1912b):

> Nothing shall be admitted as fact except what can be experienced at some definite time by some experient; and for every feature of fact ever so experienced, a definite place must be found somewhere in the final system of reality. In other words: Everything real must be experiencable somewhere, and everything experienced must somewhere be real.

minded' position.[20] He rejected it because he found the tough-minded world-view too bleak. He thought that hard phenomenalism leaves no room for truths about religion and 'human value', since if there are such truths they are non-phenomenal.[21] These criticisms of hard phenomenalism are reminiscent of his attack on naturalism in *The Varieties of Religious Experience*,[22] in which he had argued that religious belief is necessary if one is to cope with suffering without simply ignoring it. At the same time, James had no time for the 'pure but unreal' theories of 'tender-minded' philosophers (i.e. the absolute idealists). James claimed that the tender-minded have a 'feeble grasp of reality'; in consequence, he said, their theories are not in touch with our 'concrete joys and sorrows'. He summarised the problem as follows:

> You want a system that will combine both things, the scientific loyalty to facts and willingness to take account of them, the spirit of adaption and accommodation, in short, but also the old confidence in human values and the resultant spontaneity, whether of the religious or of the romantic type. And this is then your dilemma: you find the two parts of your quaesitum hopelessly separated. You find empiricism with in-humanism and irreligion; or else you find rationalistic philosophy that indeed may call itself religious, but that keeps out of all definite touch with concrete facts and joys and sorrows.[23]

James then offered 'this oddly named thing Pragmatism as a philosophy which can satisfy both kinds of demand.' The basic idea behind James's solution was that our non-phenomenal beliefs are justified by their utility. His take on this is best

---

[20]To see that when James talked about 'tough-minded philosophers' he had in mind people who accepted positions like hard phenomenalism, consider this passage from lecture VII:

> The tough-minded are the men whose alpha and omega are facts. Behind the bare phenomenal facts, as my tough-minded old friend Chauncey Wright, the great Harvard empiricist of my youth, used to say, there is nothing.

[21] I'm simplifying somewhat here because James did think that some truths about value are phenomenal, as he explains in James (1905). However, James seems to have thought that many truths about value are non-phenomenal.

[22]See James (1902), particularly the 'frozen lake' metaphor in 'The Sick Soul'.

[23]This comes from the first lecture.

understood by looking at his comments on developmental psychology. James thought that very young babies have only phenomenal concepts. For example, while a baby can identify individual sensations in its experiential input, he lacks an ontology of persisting physical things. He can see a sequence of yellow crescent-shaped colour patches, but not a banana. As James put it:

> A baby's rattle drops out of his hand, and it has 'gone out' for him, as a candle-flame goes out; and it comes back, when you replace it in his hand, as the flame comes back when relit. The idea of its being a 'thing,' whose permanent existence by itself he might interpolate between its successive apparitions has evidently not occurred to him.[24]

As a baby matures, it learns to think of sequences of similar sensations as sensations of the same persisting physical object.

The growing baby must extend its 'conceptual system' in other ways too. James thought that some spatial and temporal concepts are phenomenal: specifically, he claimed that simultaneity, adjacency and 'time interval' are given in experience.[25] However, beyond this a baby has no notion of space or time. According to James, a baby learns to think as a substantivalist as he grows up. He comes to regard 'cosmic space' as a receptacle which persists as physical objects move through it, and learns to think of as events as distributed along a time-line. He also finds ways of classifying physical things, which allow him to generalize about how things of different kinds behave.

The point of all of this, for James, is to systematize or organize the underlying phenomenal truths: to '[straighten] the tangle of our experience's immediate flux and sensible variety'[26] as he put it. Since the new-born baby has no way of organizing

---

[24]This comes from the fifth lecture.

[25]See James (1904a).

[26] This comes from the fifth lecture.

its sensory input, he experiences the world as a 'blooming buzzing confusion',[27] a chaotic tangle of sensations. The non-phenomenal concepts that he acquires as he grows up allow him to understand patterns in his experiences, so that his environment no longer seems to be just a random jumble of miscellaneous sensations.

James thought of science in much the same way. The job of the scientist, he thought, is to devise theories to further systematize the data of experience. This is a pleasurable thing to do in itself, James thought, and also allows us to make predictions more accurately and efficiently:

> Our pleasure at finding that a chaos of facts is the expression of a single underlying fact is like the relief of the musician at resolving a confused mass of sound into melodic or harmonic order. The simplified result is handled with far less mental effort than the original data; and a philosophic conception of nature is thus in no metaphorical sense a labor-saving contrivance. [For example,] [w]ho does not feel the charm of thinking that the moon and the apple are, as far as their relation to the earth goes, identical[?][28]

Here's the picture so far. James believed that we live in a 'world of pure experience', a world which consists of sensations. What's more, he thought that these sensations are (in typical cases) mind-independent; they are 'things-in-themselves', 'found, not manufactured'. He thought that perception gives us direct access to this mind-independent world of sensations, and that from a very early age we acquire from perception concepts necessary to characterise these sensations and their relations to each other. James thought that without further, non-phenomenal concepts the sensations will appear to us to be a disordered mess, a 'blooming, buzzing confusion'. So as a child grows up it extends its conceptual system with non-phenomenal concepts, which it uses to systematise or organise the phenomenal truths. James hoped that he could justify our non-phenomenal beliefs by appeal to their utility in systematising

[27]This comes from Chapter XIII of the first volume of James (1890).

[28]From James (1879a).

the underlying the phenomenal truths, so he could avoid hard phenomenalism. In particular, if he could convincingly argue that religious, moral and aesthetic claims play a pragmatically valuable role in our system of belief, he could use this to defend religion and 'human values'. At the same time, he maintained his basically empiricist or phenomenalist orientation, and in particular the claim that perception gives us direct access to a mind-independent world of sensations. This separated his position from the airy fantasies of the tender-minded absolute idealists. He could maintain 'religion' and 'human values' without losing his 'scientific loyalty to the facts'. That was the idea, anyway.

1.5 James's theory of truth

Having gone through some of the background, we are in a position to understand James's theory of truth. I will explain the theory first, and then go back and present textual evidence for my interpretation.

I'll use the term 'pragmatic virtue' for those characteristics of a theory that are desirable, for pragmatic reasons: so elegance, predictive power, computational efficiency etc. are pragmatic virtues. Using this terminology, we can state one of the conclusions of the last section as follows: for James, the aim of inquiry is to develop a pragmatically virtuous systematisation of the phenomenal truths. Given that truth is usually said to be the goal of inquiry, this suggests the following interpretation of James's theory of truth:

> *James's Theory of Truth*
>
> The truths are precisely the elements of the most pragmatically virtuous superset of the set of phenomenal truths.

I will now look at the text,[29] and argue that this was indeed James's view.

James began his discussion of truth by distinguishing 'relative' from 'absolute' truth. Since James also called relative truths 'half' truths, it seems that he identified truth simpliciter with absolute rather than relative truth. A relatively true theory, on the other hand, is a theory that is the best currently available approximation to the truth. One of James's examples is that the Copernican theory of planetary motion was relatively true back in the middle of the sixteenth century.

He then described some features of a theory which count towards its relative truth. He claimed that a relatively true theory allows one to make predictions about future phenomenal truths, and to do so efficiently. He also said that a relatively true theory must be consistent with 'relations between purely mental ideas' (in something like Hume's sense). Importantly, he added that a relatively true theory 'takes account of the whole body of other truths already in our possession.' Presumably he meant that a relatively true theory is obtained by making small changes to one's previous relatively true theory.

Having talked about relative truth, James then explained what absolute truth is as follows:

> The 'absolutely' true, meaning what no farther experience will ever alter, is that ideal vanishing point towards which we imagine that all our temporary truths will some day converge.

It's tempting to read James as endorsing the Peircean claim that a statement is true just in case it would be believed at an imaginary time in the future when inquiry has finished. I've already explained the obvious and decisive objection to this theory. But now we've put the quote in context I think we can see more clearly what James meant. The relative truths, for a given person, at a given time, systematise the

[29]The crucial lecture is the sixth.

phenomenal truths that she knows by direct experience. Since they were obtained by making small changes to the theories of past researchers, they also to a large extent systematise the observations of her predecessors. As time passes, the number of phenomenal truths known by direct experience increases, and the relatively true theory of the time systematises an ever larger body of such truths. Over time, then, the relatively true theory approaches the absolutely true theory, which is the theory which does a maximally good job of systematising all the phenomenal truths. This supports my interpretation of James's theory, which I now repeat:

> *James's Theory of Truth*
>
> The truths are precisely the elements of the most pragmatically virtuous superset of the set of phenomenal truths.

I will use the term 'best theory' for the maximally virtuous superset of the set of phenomenal truths. This allows me to summarise James's theory, as I interpret it, by saying that, according to James, the truths are precisely the elements of the best theory.

1.6 James's pluralism about truth

It might be asked, 'why THE most pragmatically virtuous superset'? Why could there not be a tie for the title 'best theory'? It turns out that James did suspect that there are ties. In the fifth lecture of *Pragmatism*, James discussed a tension between a 'common sense' description of the world and a scientific one. According to the common sense description, enduring physical objects have more or less the characteristics they appear to have; according to the scientific description, physical things are not at all as they appear, being made of invisible particles and lacking secondary qualities. James did not rule out the possibility that one of these two theories might turn out to be just false (in particular, he mentioned with sympathy

the contemporary rise in instrumentalism about scientific unobservables); however, he seemed to think it at least possible that the common sense and scientific theories are best regarded as alternative conceptual systems, to be accepted separately without being combined. He suggested that 'the conflict of these so widely differing systems obliges us to overhaul the very notion of truth, for at present we have no definite notion of what that word might mean'. 'May there not after all be a possible ambiguity in truth?', he asked. James seems to have been proposing that there is more than one sort of truth, and that the word 'true' is correspondingly ambiguous between them. This interpretation is also suggested by the following passage, from the beginning of the seventh lecture:

> What hardens the heart of every one I approach with [the pragmatist theory of truth] is that typical idol of the tribe, the notion of the Truth, conceived of as the one answer, determinate and complete, to the one fixed enigma which the world is believed to propound. The Truth: what a perfect idol of the rationalistic mind! It never occurs to most of us even later that the question 'what is the truth?' is no real question.[30]

On the most natural reading of this passage, James was saying that there are several sorts of truth. This suggests the following amended interpretation of James's position:

> *James's Theory of Truth, Pluralist Version*
>
> Say that a 'best theory' is a maximally pragmatically virtuous superset of the set of phenomenal truths. Corresponding to each best theory, there is a variety of truth. A statement has some truth-property if and only if it is an element of one of the corresponding best theory.

There's a tension between the passages in which James implied that there are several sorts of truth, and the chapter on truth itself in which James talks (without quali-

[30]From the beginning of the seventh lecture.

fication) about 'THE 'absolutely' true'. My guess is that James wasn't sure about whether there is more than one 'best system', and so he vacillated between the theory of truth that I describe in the section five and the amended, 'pluralist' version described in this section. To keep things simple, I'll put aside the pluralist version of the theory of truth from now on.

1.7 Useless truths, useful falsehoods, and buried secrets

As I said in section one, James is often associated with the view that true beliefs are just those which it profits one to have. This position is decisively refuted by the point that there are useless truths (i.e. beliefs which are true but nevertheless not profitable) and useful falsehoods (i.e. beliefs which are false but nevertheless profitable). What I want to discuss now is whether James's theory, as interpreted in section five, is refuted by this point.

First, let's think about useless truths. I take it that one or other of these is a useless truth:

(A) There were evenly many grains of rice in my rice jar at noon yesterday.

(B) There were oddly many grains of rice in my rice jar at noon yesterday.

Let's suppose that it's (A) that's true. I assume that the best theory (i.e. the maximally pragmatically virtuous superset of the set of phenomenal truths) contains statements which determine what constitutes a grain of rice,[31] and what it is for there to be evenly or oddly many of a certain sort of thing in a jar. Together with the phenomenal truths, these statements will entail (A). Thus (A) will turn up in the

[31]Insofar as this is possible—of course there will be some vagueness here.

'best theory' even though it is not itself useful to believe—assuming that the best theory is closed under entailment.[32] According to James's theory, the useless truths are statements which, although useless themselves, are consequences of a theory which is, overall, maximally pragmatically virtuous; they are theoretical spandrels.

Now let's consider falsehoods which are useful to believe. Here an example:

> John is shy, and insecure about his appearance. When he goes on a date, John sometimes gets embarrassed and finds it difficult to sustain the conversation. He knows that the problem only arises when he suspects that he is ugly. By making himself believe that he is handsome, he can save both himself and his date from long, awkward silences. It is better for John to believe that he is handsome, even though he isn't.

Here's what a proponent of James's theory should say in response to this case. John, if he is to fool himself into believing that he is handsome, is likely to have to do some tactical 'forgetting'. He might for example have to 'forget' (at least temporarily) some frank comments made by previous dates. This reflects the fact that while it is useful for John to believe that he is handsome, this belief cannot be integrated into a system of beliefs which is, overall, pragmatically virtuous. While useful for John to believe on its own, the statement that John is handsome and rich cannot be incorporated into an absolutely true total theory: it is no part of the best systematisation of the phenomenal truths.

James's theory is also not refuted by cases of 'buried secrets'. As I said in the introduction, one version of the pragmatist theory of truth is that a belief is true if

---

[32] James should say that closure under entailment is a pragmatic virtue. This ensures that the best theory is closed under entailment.

Perhaps some will find this *ad hoc*, if so there is an alternative approach: simply modify the theory of truth to say that the truths are precisely those beliefs which are entailed by the best theory (rather than those beliefs which are elements of the best theory).

and only if it would be believed at an imaginary time in the future when inquiry is finished. This is refuted by the rice jar example:

(A) There were evenly many grains of rice in my rice jar at noon yesterday.

(B) There were oddly many grains of rice in my rice jar at noon yesterday.

One or other of these is true, but since I ate some of the rice from my jar last night, nobody will ever know which. The information is lost forever. Were inquiry to finish, nobody would have an opinion on the topic. Examples like this do not threaten James's theory, on my interpretation. The fact that we will never know which of the two statements is in the best theory is not damaging to James's theory: this just shows that we are unable to know what the best theory is, because our knowledge of the phenomenal truths is limited (remember that for James some sensations will never be 'had' by anyone).

1.8 James on the plasticity of the world

As I've said, James thought that the phenomenal truths are independent of us in typical cases; we have no influence over what sensations there are, or what characteristics these sensations have, outside the small corner of space-time over which we have causal influence in the normal way. James did not have the same 'realist' attitude towards non-phenomenal truths. On the contrary, he insisted that truth is a 'man-made product', and connected this with his claim that the world is 'plastic' rather than 'ready-made'. In this section, I'll discuss these claims. James was at times quite lyrical on this point:

> In our cognitive as well as our active life, we are creative. We add, both to the subject and to the predicate part of reality. The world stands really malleable, waiting to receive its final touches at our hands. Man engenders truths upon it. No-one can deny that such a role would add both to our dignity and to

> our responsibility as thinkers. To some of us it proves a most inspiring notion. Signor Papini, the leader of Italian pragmatism, grows fairly dithyrambic over the view that it opens of man's divinely creative functions. The import of the difference between pragmatism and rationalism is now in sight throughout its whole extent. The essential contrast is that for rationalism reality is ready-made and complete for all eternity, while for pragmatism it is still in the making, and awaits part of its complexion from the future.[33]

In this passage, James is most naturally read as claiming that we exert causal influence over the non-phenomenal truths. The idea, roughly, seems to have been that the phenomenal facts are largely causally independent of us, but we then construct the non-phenomenal facts, adding them to the already existing phenomenal facts. In the final sentence of the quoted passage, James implies that this goes for non-phenomenal facts in the past, so James seems to have been committed to there being cases of backwards causation. In section 8.1, I'll explain why James should not have made these claims. I'll look at another interpretation of these passages in 8.2.

1.8.1 Do we have divinely creative functions? (Part 1)

Before giving my own reasons for saying that James should not have implied that we have causal influence over the past, I should briefly mention a bad reason for rejecting this part of James's thought:

> If we had the 'divinely creative functions' that James attributes to us, then we could create a pair of handcuffs so as to retroactively tie the hands of John Wilkes Booth, thereby preventing the assassination of Lincoln. But obviously we can't do this, so James was wrong.

This objection to James's position is mistaken; it is true that on James's view, we have the power to create past objects, but our power to do so is constrained by the

[33]This is from lecture VII.

past phenomenal facts (which are metaphysically independent of us). We can only create a pair of handcuffs where there is an appropriate sequence of handcuff-shaped sensations. So James's view does not imply that we have the power retroactively to prevent the assassination of Lincoln in this way.

But there is a better reason for rejecting James's claims about the plasticity of the world. Consider again James's theory of truth:

> *James's Theory of Truth*
>
> The truths are precisely the elements of the most pragmatically virtuous superset of the set of phenomenal truths.

This claim is supposed to apply across the board. In particular, it is supposed to apply to James's pragmatism itself. So on James's view, we should modify our theory of the world to allow for this sort of backwards causation only if doing so gives us a better systematisation of the phenomenal truths. However, as far as I can tell, our total theory does not become any better pragmatically if we modify it in this way. On the contrary, the attempt to do so seems to lead to a big mess. So James should not have said that we create the non-phenomenal things.

1.8.2 Do we have divinely creative functions? (Part 2)

But perhaps I'm being unfair. Perhaps James should not be read as saying that the non-phenomenal is causally mind-dependent. Perhaps he meant that the non-phenomenal is grounded in us. On this reading, James's talk about the world being 'malleable' and 'plastic' was colourful metaphor. What James meant was not that, for example, we cause snow to be white by including the belief that snow is white in our best theory, but rather that snow is white in virtue of our best theory's say-so.

This might seem to be the more charitable reading. The claim that we have causal influence over the past seems crazy, but the claim that the past is grounded in

the present might seem to be an attractive way of developing widely felt presentist intuitions. Even so, my opinion is that James's claims about the plasticity of the world are no more viable on the 'grounding interpretation' than they are on the 'causal interpretation'. Before explaining why, I'll need to describe the view in question more carefully, and do that I'll need some notation. Grounding is often taken to be a relation between facts. However, since James didn't talk about facts[34] we should avoid appeal to them here. Instead I'll borrow a trick from Kit Fine[35] and express grounding claims using a sentential connective '$\leftarrow$' which takes one sentence on the left and one or more right, like so:

> Europa is a moon $\leftarrow$ Europa orbits Jupiter, Jupiter is a planet

The word 'because' sometimes functions like this, as in:

> Europa is a moon because it orbits Jupiter, which is a planet.

Using this notation, and letting T be the best theory, we can express the thesis under discussion using the following schema:

> $\phi \leftarrow$ The belief that ϕ is an element of T, T is the best theory.

where for 'ϕ' one substitutes a sentence which expresses a non-phenomenal truth. The problem is immediate.[36] That T is the best theory is itself a non-phenomenal truth, and so an instance of the schema is:

> T is the best theory $\leftarrow$ The belief that T is the best theory is an element of T, T is the best theory

[34]He did use the word 'fact' a lot, but I don't think he used it in our sense.

[35]The trick comes from Fine (2001).

[36]This is a variant on the main argument in Walker (1989).

So the view in question implies that T is the best theory partly in virtue of the T's being the best theory. This seems to be an unacceptable consequence of the proposal: the 'grounds' relation is acyclic.

Tobias Wilsch pointed out to me that a pragmatist could respond by putting forward the following revised version of his schema:

> $\phi \leftarrow$ The belief that ϕ is an element of T

I don't think that this is in the spirit of James's position: it doesn't capture the idea that the non-phenomenal part of the world is dependent on our interests. Further, as Tobias also pointed out to me, the view in question leads to an ungainly regress of grounding:

$\phi \leftarrow$ The belief that ϕ is an element of T.

$\leftarrow$ The belief that [The belief that ϕ is an element of T] is an element of T.

$\leftarrow$ The belief that [The belief that [The belief that ϕ is an element of T] is ...

...

...

1.8.3 Living without divinely creative functions

So James's claims about the plasticity of the world are to be rejected, on both the 'causal' and 'grounding' readings. However, I don't think that this is a significant problem for James. Back in section five I offered this interpretation of James's theory of truth:

> *James's Theory of Truth*
>
> The truths are precisely the elements of the most pragmatically virtuous superset of the set of phenomenal truths.

James liked this theory because it implied that non-phenomenal statements can be true so long as they are elements of the best theory. In particular, James thought that some religious beliefs and beliefs about 'human values' will make it into the best theory, so on his view these beliefs are in fact true. At the same time, James thought that by maintaining a weakened form of phenomenalism, he was not losing the 'scientific loyalty to the facts' that he so valued.

Crucially, none of this depends on his peculiar claims about the non-phenomenal facts being mind-dependent. These claims simply don't do any work in James's theory—they are 'theoretical danglers'. They can be omitted with loss.

1.9 The prospects for the pragmatist theory of truth

I've now rebutted some of the standard objections to James's theory of truth. The theory, I've claimed, is not refuted by the fact that there are useless truths, or by the fact that it is sometimes useful to believe falsehoods. It is also not refuted by the fact that there are 'buried secrets'. I do think that James's claims about the plasticity of the world and our 'divinely creative functions' are to be rejected, but as I've said I don't think these claims were important to the theory anyway.

Even so, I don't expect anyone today to find the theory attractive. First of all, James was trying to solve a problem that few will now take seriously. Nobody today is losing sleep over the challenge of reconciling hard phenomenalism with Christianity, as James did. What is more, the theory itself has a phenomenalist orientation that nobody will now find attractive.

However, it does seem to me that James's problem has analogues in contemporary philosophy. For example, many philosophers accept physicalism, but at the same time have beliefs which are hard to square with physicalism: beliefs about morality, the

mental, necessity, mathematics and so on. A lot of work in contemporary philosophy is devoted to the attempt to resolve this tension.

You don't have to be a physicalist to face this sort of problem. Chalmers, for example, would say that phenomenal truths have the same kind of status as physical ones—they too are 'basic', as I will put it.[37] He still faces the problem of squaring his beliefs about what is 'basic' with his beliefs about non-basic matters—morality, perhaps, or modality, or numbers. In general, the problem is that of reconciling one's beliefs about what is basic with the rest of one's beliefs. Now it seems to me that a Jamesian theory of truth is attractive here:

> *James's Theory of Truth, Revised Version*
>
> The truths are precisely the elements of the most pragmatically virtuous superset of the set of the basic truths.

A physicalist, for example, could say that the truths are precisely the elements of the most pragmatically virtuous superset of the set of physical truths. This is recognisably a physicalist position. At the same time, the physicalist-pragmatist could defend his beliefs about mind, number, morality and so on by arguing that they are elements of the 'best theory'. To put it rather grandly, you could think of this as a way of resolving the tension between the manifest and scientific images.[38]

I don't have a decisive objection to this proposal, but nevertheless I am suspicious. I will finish by outlining two reasons for my suspicion.

---

[37]See Chalmers (1996).

[38]The terms 'scientific image' and 'manifest image' come from Sellars (1963).

1.9.1 First reason: the children and animals objection

As I observed back in section five, James's theory of the goals of inquiry coheres with his theory of truth:

> The theory of inquiry: The goal of inquiry is to produce pragmatically virtuous systematisations of phenomenal truths.
>
> The theory of truth: The truths are precisely the elements of the most pragmatically virtuous systematisation of all the phenomenal truths.

These two theories fit together elegantly; they make good on the common idea that 'truth is the goal of inquiry'. The proposed updated version of James's theory of truth does not have this attractive feature. It is not plausible, from a contemporary perspective, that in general the goal of inquiry is to produce pragmatically virtuous systematisations of the basic truths. Perhaps scientists, philosophers are other intellectuals are interested in the systematisation of the basic truths—but this is unusual. Animals and children, for example, are not interested in systematising the basic truths.

This means that the proposed updated version of James's theory of truth is not as well-motivated as the original version.

1.9.2 Second reason: the problem of pragmatically deficient concepts

The second objection is a variant on the 'useless truths' objection, considered back in section one. I hereby introduce the concept sib by stipulating that something is a sib if and only if it is either an elephant, or a triangular number, or a country other than Nepal. Given this stipulation, it is true that Italy is a sib. However, since the concept sib is a useless, this truth won't appear in the best theory. So it is a counterexample to the revised version of the pragmatist theory of truth.

One could save the theory in a rather *ad hoc* manner by arguing that to contain analytic truths is a pragmatic virtue of a theory.[39] This would ensure that:

All countries other than Nepal are sibs

gets a place in the best theory. Given that:

Italy is a country and Italy is not Nepal.

is also in the best theory, and given that the best theory is closed under entailment, it will follow that the belief that Italy is a sib is an element of the best theory after all. As I say, this strikes me as an *ad hoc* manoeuvre, but I need not dwell on the point because the 'problem of pragmatically defective concepts' arises in other cases, which can't be avoided in the manner just discussed. Consider for example the concept *pigeon*. This is, while not useless, a pragmatically deficient concept: the distinction made in English between doves and pigeons serves no function. No best theory would bother with this distinction. But it is surely true that there are many pigeons in Trafalgar Square. Since the concept *pigeon*, unlike *sib*, doesn't have a definition, this counterexample cannot be explained away in the manner just discussed.

1.10 Postscript: James, Russell, Carnap, and Quine

Russell's scorn for James's pragmatism is well known. Russell[40] and Moore[41] made objections to James's view that have subsequently become canonical.[42] Russell was

[39]Indeed, James seems to have held something like this. See his discussion of 'relation between purely mental ideas', in lecture VI.

[40]See Russell (1992) and Russell (1909).

[41]Moore (1922).

[42]See section one.

quite passionate in his criticism of pragmatist claims about the plasticity of the world:

> In all this I feel a grave danger, the danger of what might be called cosmic impiety. The concept of 'truth' as something dependent upon facts largely outside human control has been one of the ways in which philosophy hitherto has inculcated the necessary element of humility. When this check upon pride is removed, a further step is taken on the road to a certain kind of madness—the intoxication of power which invaded philosophy with Fichte, and to which modern men, whether philosophers or not, are prone. I am persuaded that this intoxication is the greatest danger of our time, and that any philosophy which, however unintentionally, contributes to it is increasing the danger of vast social disaster.[43]

While it strikes me as extremely unlikely that James's publications have made some 'vast social disaster' non-negligibly more probable, I agree with Russell that this part of James's doctrine is to be rejected. However, as I have said, I think that James's 'impious' claims can be excised from his theory will little change to the rest.

Russell's attitude to James's pure experience theory was very different. Russell very much admired the pure experience theory, and indeed he regarded his own neutral monism as a development of the pure experience theory.[44]

Russell argued, for example in 'The Relation of Sense-Data to Physics',[45] that truths not expressible in phenomenal terms would be unknowable for us, and so the knowable portion of the world is completely describable in phenomenal terms. Like James, he recognised that a merely phenomenal language would be completely impractical for everyday life and for science, and so we have a practical need for other terminology. Russell departed from James, however, by suggesting that the words in our language which are neither phenomenal nor logical are explicitly definable using the phenomenal ones. Just as, in real arithmetic, one defines all one's terms

---

[43]Russell (1946). The quote is taken from the end of the chapter on Dewey.

[44]See Putnam (1992) for discussion.

[45]Russell (1914).

('differentiable', 'stationary point', etc.) using a small number of primitives ('zero', 'is greater than' etc.), so—Russell thought—one could define 'horse', 'electron', 'Everest' and all the rest in phenomenal terms. It was the project of providing such definitions, the project of 'analytic reductionism', which Russell began in his 'The Relation of Sense Data to Physics' and which Carnap continued in the *Aufbau*.[46]

When Quine launched his famous assault on the 'Two Dogmas of Empiricism',[47] it was the Russell/Carnap theory—outlined very briefly above—that was his target. Quine rejected the Russellian or Carnapian claim that each sentence is analytically equivalent to a sentence in phenomenal language on the grounds that there is no such thing as analyticity, and in any case 'the unit of empirical significance is the whole of science'. Roughly, then, Russell modified James's theory by adding analytic reductionism, and Quine modified Russell's theory by taking the analytic reductionism out again. In consequence, the theory which Quine presented at the end of 'Two Dogmas' is similar in many respects to James's theory, presented in *Pragmatism*[48] and in the Essays in Radical Empiricism.[49] The similarity is particularly clear in this passage, which appeared in the original version of 'Two Dogmas':

> Imagine, for the sake of analogy, that we are given the rational numbers. We develop an algebraic theory for reasoning about them, but we find it inconveniently complex, because certain functions such as square root lack values for some arguments. Then it is discovered that the rules of our algebra can be much simplified by conceptually augmenting our ontology with some mythical entities, to be called 'irrational numbers'. All we continue to be really interested in, first and last, are rational numbers; but we find that we can commonly get from one law about rational numbers to another much more quickly and simply by pretending that the irrational numbers are there too.

---

[46]Carnap (1928). For an alternative reading of the *Aufbau*, see Friedman (1987).

[47]Quine (1953).

[48]James (1907).

[49]James (1912a).

> I think this a fair account of the introduction of irrational numbers and other extensions of the number system. The fact that the mythical status of irrational numbers eventually gave way to the Dedekind-Russell version of them as certain infinite classes of ratios is irrelevant to my analogy. That version is impossible anyway as long as reality is limited to the rational numbers and not extended to classes of them.
>
> Now I suggest that experience is analogous to the rational numbers and that the physical objects, in analogy to the irrational numbers, are posits which serve merely to simplify our treatment of experience. The physical objects are no more reducible to experience than the irrational numbers to rational numbers, but their incorporation into the theory enables us to get more easily from one statement about experience to another.

Setting aside the style of this passage, and the use of the mathematical analogy, it could have been written by James

Quine himself seems to have been unaware of this connection between his own views and those of James—and in fact he was rather scathing about both James and Peirce[50]—but the connection is there nonetheless. The influence of James's pragmatism on philosophy in the first half of the twentieth century was very deep.

[50]Koskinen and Philström (2006).

Chapter 2

Does the Truth have Cash Value?

2.1 Introduction

Cornelius knows that he shouldn't eat any more sticky buns; he also knows that he is likely to do so if he gets the chance. So he asks his wife to hide them. Cornelius thinks—and he may be right—that true and sufficiently specific beliefs about the location of the buns would do him no good. This shows that having an extra true belief doesn't always make a person better off. The same goes for groups of people. When things are going well, an insurance company redistributes money from the fortunate to the unfortunate—from people who need money less to people who need it more. But the system only works because nobody knows in advance who will be lucky. Obviously, the fortunate have no motivation to take part if they know who they are in advance, and the insurance company has no incentive to insure the unfortunate if it can identify them. The whole system is built on ignorance. In this case, we do better when we know less.

So it is pretty clear that there are cases in which having true beliefs makes you worse off. However, we usually think that such cases are exceptional. In typical cases, we think, ignorance is not bliss. Funny cases aside, truth is precious.

At a first pass, we think that this is a good rule to follow:

For any proposition P, try to:

- Believe P if P is true.
- Do not believe P if P is false.

This requires a small modification. There are some propositions—'irrelevant' propositions, as I call them—that are simply not important to us, whether true or false. Consider for example the proposition that the $1{,}234^{\text{th}}$ digit in the decimal expansion of $\sqrt{5}$ is a '7'. Acquiring true but irrelevant beliefs is usually a bad idea, because it's a waste of effort. I don't recommend spending a lot of time learning the decimal expansion of $\sqrt{5}$. So the truth rule needs to be modified:

[TR] For any proposition P, try to:

- Believe P if P is true and relevant.
- Do not believe P if P is false or irrelevant.

This is no iron law. It has exceptions, as we have seen. Even so, I think it's recognisably a rule that we accept, funny cases aside.

However, in the penultimate chapter of his book *The Fragmentation of Reason*, Steven Stich argues that all of this is completely wrong-headed. 'Once we have a clear view of the matter,' he says, 'most of us will not find any value, either intrinsic or instrumental, in having true beliefs'.[1]

It is important to be clear about how radical Stich's claim is. Stich is not just drawing attention to the fact that there are exceptions to the truth rule [TR]—cases in which having true (and relevant) beliefs does one no good. Nor he just claiming

[1]Stich (1990), pg. 108. From now on, I will only specify the page number when citing this book.

that the truth rule needs to be modified or qualified for some reason. He thinks that truth has no normative significance whatsoever.

2.2 Stich's views about belief and truth

I will set out Stich's argument in section three. In this section, by way of preparation, I will look at some of Stich's assumptions about what beliefs are, and about what it takes for a belief to be true.

Stich thinks that computation in general, and so mental processes in particular, involve the manipulation of symbols. In the case of human cognition, Stich supposes, the relevant system of symbols is similar in many ways to the sort of formal language we study in logic. At least, he thinks the analogy is sufficiently close that it makes sense to call this symbol system 'the 'language of thought' ('Mentalese' for short), and talk about mental 'words' and mental 'sentences' without too much distortion. Mentalese is an interpreted language: some Mentalese sentences express propositions. This implies that there is a (partial) function from the class of Mentalese sentences to the class of propositions, which maps Mentalese sentences to the propositions they express. I will call this 'the standard interpretation function'.[2,3]

To have a belief, according to Stich, is to 'is to have a token of a well-formed formula stored appropriately in one's brain'.[4] To be more precise about it, to believe a proposition P is to bear a certain relation to a token Mentalese sentence, which

---

[2]Stich also considers views on which this function maps Mentalese sentences not to propositions, but to 'content sentences, or specifications of truth conditions[, or] possible facts, or states of affairs, or subsets of the set of possible worlds' (pg. 104). As far as I can tell, it doesn't much matter for our purposes which of these alternatives one chooses, so I'll just stick to propositions.

[3] Perhaps a person's language of thought is vague, in which case its semantics is arguably best specified by a cluster of similar interpretation functions, corresponding to 'precisifications' of the language, rather than a single interpretation function. As far as I can tell, this point complicates the discussion without changing it any important way—so I put it aside.

[4]Pg. 109.

is mapped to P by the standard interpretation function. It's an empirical question what this 'certain relation' is, a question best left to cognitive scientists. But we can visualise the whole thing by supposing that inside each person's head there's a big box labeled 'Beliefs', inside which there's a pile of token Mentalese sentences. The propositions that a person believes, then, are just the values of the standard interpretation function when applied to the sentences in the box.[5]

So Stich makes some strong claims about our cognitive architecture—claims that should be assessed by scientific standards. However, it is to be emphasised that Stich offers these ideas as an account of belief *in the folk sense of that term.* He is not suggesting a scientific substitute for the folk concept; he is attempting to describe in scientific terms the very things which the folk call 'beliefs'. 'My scepticism about the value of truth,' Stich explains, 'is restricted to accounts that assign truth conditions largely compatible with commonsense intuition'.[6]

Now the standard interpretation function is not the only function which maps Mentalese sentences to propositions—there are many others. We can call the other ones 'alternative' interpretation functions. It's natural to ask: what identifies the standard interpretation as such? Why is *it* the standard one?

Stich offers a sketch of an answer to this question. He imagines that the language of thought contains proper names and predicates, and that the proper names refer to individuals while the predicates refer to properties or relations. Stich supposes that some appropriately modified version of Kripke's causal theory of reference[7] will be sufficient to explain which Mentalese terms refer to what. The standard interpretation will then map a sentence of the form $\ulcorner Ra_1, \ldots a_n \urcorner$ to the proposition that the objects referred to by $a_1, \ldots, a_n$ stand in the relation referred to by R.

---

[5]This piece of imagery is due to Schiffer; Stich mentions it approvingly.

[6]Pg. 106.

[7]Described, of course, in Kripke (1980).

As to more complex sentences, Stich suggests that the standard interpretation will be characterised recursively, with the recursive clauses chosen to respect the inferential roles of the Mentalese connectives:

> it is the pattern of interactions among belief inscriptions which is essential. If the patterns of interactions manifested by sentence of the form 'P*Q' approximates the pattern one would expect if '*' were the symbol for the material conditional, then '*' *is* the symbol for the material conditional. Similarly for the rest of the connectives and quantifiers.[8]

Obviously, there are lot of details to be worked out here: Stich's theory is far from complete, as he is well aware. But enough detail has been given to support the subsequent argument, which is what matters.

2.3 The gist of Stich's argument

In this section, I present the basic outline of Stich's main argument. I will present it again, more methodically, in the next section. To begin, let's look in a bit more detail at Stich's views about reference. As a refresher, here's the gist of Kripke's theory of reference for proper names in spoken languages. A name first gets its reference when a 'baptism' takes place. Someone might say:

> Let's call this baby 'Humphrey'.
>
> I name this ship 'The Pelican'.
>
> There must be a planet, 'Neptune' let's say, that's causing these perturbations in Uranus's orbit.

Some baptisms won't be quite as explicit as this, of course. Once the name has been introduced, it can then get transmitted from person to person, all the time referring

[8]Pg. 110.

to the thing specified at the original baptism. Kripke recognised that there are a lot of details missing:

> [My account of reference] has been far less specific than a real set of necessary and sufficient conditions for reference would be. Obviously the name is passed on from link to link. But of course not every sort of causal chain reaching from me to a certain man will do for me to make a reference. There may be a causal chain from our use of the term 'Santa Claus' to a certain historical saint, but still the children, when they use this, by this time probably don't refer to that saint. So other conditions must be satisfied in order to make this into a really rigorous theory of reference. I don't know that I'm going to do this because, first, I'm sort of too lazy at the moment; secondly, rather than giving a set of necessary and sufficient conditions which will work for a term like reference, I want to present just a better picture than the picture presented by the received views.[9]

So Kripke envisaged, but did not attempt to specify in detail, a theory of this form:

> Token name n refers to thing x just in case there is an appropriate causal chain connecting n to a baptism at which x was dubbed with a type-identical name.

To get a thorough version of the theory, one would have to explain what an 'appropriate causal chain' is, and what it is for an event to be a baptism at which object x is given name n.

Kripke intended all of this as an account of reference for proper names in spoken languages. Stich supposes that the account could be adapted to provide an account of reference for Mentalese proper names and predicates. Like Kripke, Stich does not attempt to spell out the theory in any detail. According to Stich, to do so would be very difficult: the theory would have to be spectacularly complicated. Stich claims that there are many different sorts of baptism, and many different sorts of appropriate causal chain. In consequence, he argues, any well-worked out theory of reference for

[9]Kripke (1980), pg. 93.

Mentalese would have to be horrendously disjunctive. There are many very different sorts of reference-fixing causal relation, and such relations are connected only by a 'loosely knit fabric of family resemblances':[10]

> Proper names and nicknames get affixed to all sorts of things—babies, popes, battleships, breakfast cereals, islands, wars, and tyrants, to mention just a few—and the baptismal processes typically involved differ markedly from one sort of object to another. It is hard to believe that they constitute anything like a natural kind. The heterogeneity of intuitively acceptable groundings grows even more extreme when we consider the ways in which predicates come to be paired with their extensions. 'Gold', 'helium', 'asteroid', 'electron', 'kangaroo', and 'superconductivity' are, presumably, all natural kind terms, but their groundings are sure to have been very different from each other in all sorts of ways. The processes or reference-preserving transmission are comparably diverse. None of this, I hasten to add, is intended as a criticism of the causal theory as an explication of our pretheoretic views about how words in a public language or a mental language are related to what they designate. My point is simply that any plausible elaboration of [the causal theory of reference] will specify lots of allowable causal patterns. The causal chains linking my mental tokens of the names of my children to the appropriate young people are very different from the causal chain linking my token of 'Socrates' to Socrates. And both of these chains are notably different from the one linking my token of 'water' with water and from the one linking my token of 'quark' with quarks. What ties all these causal chains is that commonsense intuition counts them all as reference-fixing chains.[11]

On Stich's view, then, given an adequate definition of the 'standard' reference relation, we could tweak the definition in order to obtain definitions of other reference-like relations. Indeed, Stich argues that there is a 'bristling infinity'[12] of such alternative reference relations—we could call them 'REFERENCE*', 'REFERENCE**' and so on. Stich then points out that these reference-like relations could be used to define alternative interpretation functions. In addition to the standard interpretation function, 'I_0' say, there are a large number of alternatives: I*, I**, I*** etc.. Then just

---

[10]Pg. 114.

[11]Pp. 114-5.

[12]Pg. 119.

as a belief is said to be true if it is mapped by I_0 to a true proposition, we can say that a belief is TRUE* if it is mapped by I* to a true proposition, and TRUE** if it is mapped by I** to a true proposition, and so on.

So now the question is: Why prefer true beliefs to TRUE* ones, or TRUE** ones? What makes the standard interpretation special, from a normative point of view?

Stich's view is that the standard interpretation function has an 'idiosyncratic nature.'[13] It is a 'hodgepodge'. The only thing, Stich argues, that marks out the standard interpretation function as in any way special is that it accords with commonsense intuitions—intuitions which, Stich supposes, arise from an accumulated body of arbitrary traditions.[14] So on Stich's view we have no reason in general to prefer true beliefs to TRUE* ones, and that there is nothing special, from a normative point of view, about the standard interpretation function. Stich's argument, in brief, goes like this. We have no reason for thinking that truth is any more important, from a normative point of view, than TRUTH*, TRUTH** and TRUTH*** and so on. So the claim that we should 'pursue truth' is not well motivated.

In the next section, I will present the argument step-by-step. Before then, I need to get an objection out of the way. The sentence 'My love gets bigger by the day' is ambiguous. On one reading, it is declaration of ever increasing love; on the other, it is a declaration of love for one who is ever increasing. Sometimes 'love' refers to the person loved, sometimes to the state of the lover. There is a similar, though less conspicuous, ambiguity in 'belief'. 'John's belief' can be used to refer to the proposition that John believes; it can equally be used to refer to John's state of belief.[15] To avoid equivocation, I will distinguish the 'belief-state' from the

---

[13]Pg. 119.

[14]Stich also mentions that the possibility that the intuitions have a genetic basis, but says that he thinks this 'vastly less likely' (pg. 120).

[15]At least, the word 'belief' in the philosophical lexicon is ambiguous in this way. I am not sure if this ambiguity exists in normal English.

'proposition believed'. The arguments of interpretation functions are belief-states, the values of such functions are the propositions believed.

Now it is sometimes argued that 'true', in its proper sense, applies to propositions only; it applies to belief-states only in a derivative and perhaps illegitimate sense. William Alston used this to object to Stich.[16] He claimed that TRUTH*, TRUTH** and so on are not genuinely truth-like properties at all: truth is a property of propositions, TRUTH* is a property of belief-states. But it's fairly easy to modify Stich's argument to avoid this objection, as Alvin Goldman has pointed out.[17] Rather than using the alternative interpretation functions—I*, I** and so on—to define alternative truth-like properties, we can use them to define alternative belief-like relations. A person believes a proposition P, on Stich's view, if the person bears an appropriate relation to a Mentalese sentence S, and $I_0(S)$=P. Similarly, we can say that a person BELIEVES* a proposition P if the person bears the appropriate relation to a Mentalese sentence S, where I*(S)=P. 'BELIEVES**', 'BELIEVES***' and so on can be defined similarly. We can then ask why we aim to believe truths, when we could aim to BELIEVE* truths, or BELIEVE** truths.

2.4 A reconstruction of the argument

In this section, I will present a version of Stich's argument in premise conclusion form. Here's the truth rule again:

[TR] For any proposition P, try to:

- Believe P if P is true and relevant.
- Do not believe P if P is false or irrelevant.

[16] In Alston (1996).

[17] See Goldman (1991).

And here are some alternative rules:

[TR*] For any proposition P, try to:

- Believe P if P is TRUE* and relevant.
- Do not believe P if P is FALSE* or irrelevant.

[TR**] For any proposition P, try to:

- Believe P if P is TRUE** and relevant.
- Do not believe P if P is FALSE** or irrelevant.

Stich's conclusion is that we have no good answer to the question, 'Why follow the truth rule?'

Here's the argument:

(1) We have no reason to think that there is any more intrinsic value to having true beliefs than there is to having true BELIEFS*, or true BELIEFS**, etc.. (Premise)

(2) We have no reason to think that following [TR] is any better instrumentally than following [TR*], or [TR**] etc.. (Premise)

(3) Conservatism alone does not provide a sufficient reason for following [TR] rather than [TR*], [TR**] etc.. (Premise)

(4) We have no account of why [TR] is a good rule to follow, when compared to the alternatives [TR*], [TR**] and so on. (From 1, 2, 3)

(5) A good answer to the question 'Why follow [TR]?' would explain why [TR] is a good rule to follow, when compared to the alternatives [TR*], [TR*], [TR**] and so on. (Premise)

(6) We have no good answer to the question, 'Why follow [TR]?' (From 4, 5)

As I've reconstructed the argument, its conclusion is not as dramatic as the advertised claim that truth has no normative significance. Even so, it is extremely counterintuitive. Notice that a parallel argument could be used against someone who recommended an alternative truth rule—perhaps a rule that incorporates some extra qualification. Now let's look at the argument step by step.

Start with premise (1). Stich could be read as arguing that because the standard interpretation function is a 'hodgepodge', there is no intrinsic value to having true beliefs. This would not be a good argument: it could reasonably be responded that true beliefs are valuable even though they are miscellaneous. Why, after all, should the good things constitute a 'natural kind'? The position can be illustrated by looking at other things which have been claimed to be intrinsically valuable. Consider, for example, Moore's views on the matter:

> By far the most valuable things, which we know or can imagine, are certain states of consciousness, which may roughly be described as the pleasures of human intercourse and the enjoyment of beautiful objects. No one, probably, who has asked himself the question, has ever doubted that personal affection and the appreciation of what is beautiful in Art of Nature, are good in themselves ... [18]

Plausibly the class of appreciations-of-beautiful-things is hodgepdogey. It doesn't seem to follow that Moore is wrong to say that the appreciation of what is beautiful is good in itself. For a second example, consider utilitarianism. Utilitarians claim that only pleasures are intrinsically valuable. It might be argued that the pleasures are miscellaneous. Certainly, they seem to be very different from one another phenomenologically. The pleasure of getting into a hot bath on a cold day doesn't feel much like the pleasure of eating chocolate. Even if this is right, it doesn't seem to follow that utilitarianism is false.

So Stich doesn't seem to have a good case for the conclusion that there is no intrinsic value to having true beliefs. However, I think that Stich does have a good

case for claim (1). If Stich's views about belief and truth are correct, it is hard to see how to make a *positive case* for the claim that truth is valuable.

Now let's look at the second premise: does having true beliefs help us to achieve other important goals? Stich's answer to this question is simply that we have no reason at all to think that true beliefs are any more useful instrumentally than TRUE* ones, or TRUE** ones. Stich imagines an opponent claiming that a preference for true beliefs (over TRUE* beliefs etc.) works out better 'in the long run',[19] and comments simply:

> Well perhaps it does. But to show this requires an argument, and as far as I know, no one has any inkling of how that argument would go.[20]

The rest of the argument requires less discussion. I take it that few people will think it a good idea to follow [TR] solely out of respect for tradition, so premise (3) seems fairly secure. If true beliefs are neither intrinsically or instrumentally valuable, it would seem that only conservatism could motivate the following of [TR], so (1), (2) and (3) together do seem to imply (4). Premise (5) is an instance of the more general principle that a good reason for doing something is always a reason for doing it rather than the available alternatives. The inference to (6) from (4) and (5) seems obviously sound.

2.5 Stich's pragmatism

Here is an obvious question: if we are not to aim at having true beliefs, what should we aim at, when forming beliefs? Stich devotes a whole chapter to this question.[21] I

[19] Pg. 123.

[20] Pp. 123-4.

[21] Chapter 6.

will, however, just quote one passage which should be sufficient to explain the basics of Stich's proposal:

> But if truth is not to be the standard in epistemology, what is? The answer that I favor is one that plays a central role in the pragmatist tradition. For pragmatists, there are no special cognitive or epistemological values. There are just values. Reasoning, inquiry and cognition are viewed as tools that we use in an effort to achieve what we value. And like any other tools, they are to be assessed by determining how good a job they do at achieving what we value. So on the pragmatist view, the good cognitive strategies for a person to use are those that are likely to lead to the states of affairs that he or she finds intrinsically valuable.[22]

Consider, for example, the beliefs that are causally involved in one's egg-scrambling. When making decisions about what to believe here, one should aim to form and retain beliefs that are helpful in producing better scrambled eggs, and in other similarly practical ways.

2.6 Some responses to the argument

2.6.1 First response

Stich assumes that it is rational for us to intrinsically value true beliefs only if we have some reason for valuing truth above (for example) TRUTH. But this assumption is quite unwarranted.*

[22]Stich (1993).

Consider my friend Alan, who intrinsically values enlarging his collection of:

- *stamps;*
- *green model trains;*
- *glass eighteenth century bottles;*
- *books about flamingos; and*
- *fossilised ammonites.*

Now Alan is quite aware that his collection is 'disjunctive', but he's completely unfazed by this—and I can't see any reason for thinking that Alan's (admittedly somewhat strange) preferences are evidence of any kind of irrationality. Now I think we could say the same thing about our preference for truth. Perhaps Stich is right that there is something arbitrary about our preference for true beliefs, but this doesn't make the preference irrational.

I am sure that Stich is aware that this sort of position is available to his opponents,[23] and as far as I can see Stich has no argument against views of this kind.

However, it is important to notice that this response to Stich is *highly* concessive: we do not normally think that our preference for true beliefs is arbitrary in this way. One way to see this is to note that we not only value true beliefs for ourselves, we encourage other people to do the same, and look down on people who do not 'respect' the truth. It is hard to square this practice with the suggestion that our preference for truth is arbitrary.

---

[23] I think that this is what he has in mind when he says on pg. 120, 'nothing that I have said comes even close to a knockdown arguments against according intrinsic value to having true beliefs.'

2.6.2 Second response

We know from everyday experience that it is instrumentally valuable to have true belief-states. I could tell you lots of stories to illustrate this.

I am very sympathetic with this appeal to 'casual empiricism'. However, on its own I find this unsatisfying. If [TR] is a better rule to follow that the alternatives, we should like to know why.

2.6.3 Third response

Stich may be right that we have no reason to value true belief-states over (for example) TRUE belief-states. However, perhaps tragically, we cannot help doing so. Whatever our assessment of Stich's argument, we will continue to value truth.*

I don't see any reason for thinking that this is so. It is of course often difficult to overcome long-held habits, and so it might well be true that it would be difficult for us to stop attempting to value truth—but I don't see any good reason for thinking that the attempt would surely meet with failure.[24]

2.6.4 Fourth response

Stich supposes that the folk have intuitions about the reference relation, construed as a relation between Mentalese 'words' and the objects of our thoughts. But this is false—the folk have no such intuitions. The relevant concept of reference is a theoretical concept from cognitive psychology, a subject in which the folk have, for the most part, little interest.[25]

Stich does not assume that the folk have intuitions about reference, so understood. He does assume that the folk have intuitions about the truth or falsity of belief-states.

---

[24]See the interchange between Stich and Bishop in Bishop and Murphy (2009).

[25]See Bach (1992).

The relevant concept of reference is introduced as part of a theoretical account of truth, in the folk sense of that term.

2.6.5 Fifth response

There is, after all, an important difference between truth and TRUTH: truth has 'the disquotational property', while TRUTH* does not.*

We should begin by saying something about what could be meant by saying that truth has 'the disquotational property'. When we want to talk about specific propositions, we usually use 'that-clauses'. We talk, for example, about 'the proposition that snow is white' and 'the proposition that absence makes the heart grow fonder'. These labels are used in such a way that, apart from perhaps a few pathological cases, instances of the following schema are true:

The proposition that S is true iff S.

Let's call the 'non-pathological' sentences 'safe'. We can then argue as follows. I believe the proposition that snow is white. Call the relevant Mentalese sentence '*S*'. Then:

S is true iff I_0(S) is true.
iff the proposition that snow is white is true.
iff snow is white.

A few pathological cases aside, we accept instances of this schema (where for 'S' one substitutes a declarative sentence):

If an agent believes the proposition that S, her belief-state is true iff S.

We could summarise this by saying that truth has 'the disquotational property'. Now TRUTH* does not have the disquotational property, in this sense. That is, there are truth-apt sentences S which do not yield a true sentence when substituted into:

> If an agent believes the proposition that S, then her belief-state is TRUE* if and only if S.

This is, indeed, a difference between truth and TRUTH*. Does this show that there is something wrong with Stich's argument? I think not. The important point to notice here is instances of this schema are true:

> If an agent BELIEVES* the proposition that S, her belief-state is TRUE* if and only if S.

This can be shown using the same argument used above to show that truth has the disquotational property, *mutatis mutandis*. We could say that while TRUTH* lacks the disquotational property, it has the DISQUOTATIONAL* property. Truth, of course, lacks the DISQUOTATIONAL* property. In the absence of some reason for thinking that the disquotational property is important in a way that the DISQUOTATIONAL* property is not, it is unproblematic for Stich that truth and TRUTH* differ in this way.

2.6.6 Sixth response

The norm of belief-formation under discussion really has little to do with truth. The norm is rather the conjunction of all instances of the following: Believe P only if P. Truth has nothing to do with it.[26]

[26]See Harman (1991).

Perhaps so. But Stich of course will ask, why this schema rather than one of these?

BELIEVE* P only if P.

BELIEVE** P only if P.

BELIEVE*** P only if P.

...

...

2.6.7 Seventh Response

Our preference for true beliefs just follows from a norm of consistency. One can't consistently believe both P and that P is not true. Hence, to be consistent, we must refrain from believing things that we believe not to be true.

This has a superficial ring of plausibility to it, but one can see that it is specious by asking 'What sort of consistency is at issue here?'. I take it that the beliefs are 'consistent' in the relevant sense if they could all be true. Since a belief that P and a believe that P is not true cannot both be true, they are inconsistent. Analogously, we could define 'CONSISTENT*' by saying that true beliefs are CONSISTENT* only if they could both be TRUE*. The question is then: why should we care about consistency rather than CONSISTENCY*?

Secondly, a consistency norm of the proposed kind imposes only a synchronic constraint on rationality: it would tell us not now to believe things which we currently take to be false. It would not tell us not to form beliefs in the future which we currently take to be false.

2.7 Goldman's 'tu quoque'

In the last section I looked at seven objections to Stich's argument, none of which is successful. In this section I look at a sixth objection, due to Alvin Goldman.[27] This objection is successful, once a small correction has been made.

> If [Stich's Argument] were right, however, a precisely analogous problem would apply to [Stich's version of] pragmatism. There would be equally legitimate alternatives for the mapping function from desire-states to conditions of fulfillment (or satisfaction), and we would have no reason to heed the aim of ordinary desire fulfillment as compared with desire fulfillment*, fulfillment**, and so forth. Moreover, since having a reason to prefer something, according to pragmatism, depends on that thing's best satisfying one's intrinsic desires, if there are alternative accounts of satisfaction (namely satisfaction*, satisfaction**, and the like) and hence alternative accounts of what one intrinsically desires, it is impossible to have a determinate reason to prefer one thing over another. Thus, pragmatism would run afoul of precisely the same problem as the truth approach, assuming that the latter is a bona fide problem.[28]

Goldman interprets Stich as accepting a norm like this:

> [DR] For any possible belief-state B, go into that belief state if and only if doing so will help to ensure that your intrinsic desires are fulfilled.

As Goldman points out, one can construct alternatives to [DR]—[DR*], [DR**], and so on—and then argue that we have no reason for following [DR] rather one of the alternatives, using an argument that parallels Stich's own argument. Goldman's claim is that this argument stands or falls with Stich's argument against the acceptance of the truth rule [TR]—so either Stich's argument against the truth rule fails, or his own position is vulnerable to a parallel argument.

There is an imperfection in Goldman's discussion, which is easily fixed. When outlining his pragmatism, Stich does not say that we should aim to choose beliefs that

---

[27] See Goldman (1999).

[28] Goldman (1999), pp. 72-3.

help us bring about the fulfillment of our desires. He talks instead about choosing beliefs which help us secure what we intrinsically value—which is not quite the same thing. Goldman seems to have slightly misunderstood Stich's position. Stich is better interpreted as endorsing a rule like this:

> [VR] For any possible belief-state B, go into that belief state if and only if doing so will help you secure what you intrinsically value.

However, standard theories about what it is to value something imply that what one values supervenes on what one believes and desires in such a way that alternative interpretation functions will correspond to alternative, value-like relations—so one can use the interpretation function I* to define a relation VALUE*, and the interpretation function I** to define a relation VALUE** and so on. Consider, for example, the Lewisian claim that one values something if one desires to desire it, or Quinn's suggestion that to value something is to believe it to be good.[29] Even if attempts like this to reduce valuing to belief and desire do not succeed, it seems (since valuing is an intentional state) that different interpretation functions will correspond to different, valuing-like relations. So Goldman's point still applies, in a slightly modified form. Alongside the 'value rule' [VR] there is a 'VALUE* rule' [VR*] and a 'VALUE** rule [VR**] and so on. It could be argued that we have no reason for following [VR] rather than one of these other rules, and so it does indeed seem—as Goldman claims—that Stich's pragmatism is vulnerable to precisely the sort of criticism that Stich makes of the orthodox idea that we should aim to believe truths. Either Stich has been hoist with his own petard, or he doesn't have a petard, in which case he can't hoist anyone else.

---

[29]See Lewis (1989) and Quinn (1993).

Goldman's 'tu quoque' shows that something has gone wrong with Stich's discussion. However, no flaw in the original argument has been identified.

2.8 Introducing my response to the argument

In this section I want to look more critically at Stich's views about reference. Stich supposes that we can explain what it is for a Mentalese word to refer to a thing or property by using an adapted version of Kripke's causal theory of reference. I doubt that this could work. Consider, for example, the Kripkean notion of baptism. At one of Kripke's baptisms, a person identifies an object and stipulates that a certain word is to refer to the object. For example, a person identifies a baby, and declares that the baby is to be called 'Tabitha'. At a Stichean baptism, presumably, one would identify an object in thought and then carry out some kind of internal stipulation which associates the object with a Mentalese term. However, to baptise something in this way, one must already have an ability to have thoughts about it. It is a mystery, on this view, how the first baptism could ever take place. Perhaps Stich would say that one identifies the object using some kind of Mentalese indexical—but in this case, what determines what that indexical refers to?

So I don't think that Stich's theory of reference is correct.

Anticipating this sort of objection, Stich says:

> [My theory of reference] facilitated the argument by making vivid the existence of vast numbers of alternative, nonintuitively sanctioned mappings from beliefs to truth conditions or propositions. However, once the point has been made, it will stand no matter what account of the interpretation function a theorist proposes. Whatever function it is that best explicates commonsense intuition, there will be heaps of alternative functions that don't best explicate commonsense intuition. ...

> On any account of the interpretation function, to make a case for the instrumental value of true beliefs requires showing not just that they are better than false beliefs, but also that they are better than TRUE* ones, TRUE** ones, TRUE*** ones, and all the rest.[30]

I agree—merely to point out problems with Stich's theory of reference does little to undermine his argument. However, it seems to me that if Stich's theory of content determination implies that truth is of little normative significance, we can reasonably regard this as (another) problem for the theory; this motivates a search for an alternative account, one which will enable us to explain why true beliefs are of value. That's what I'll be doing over the next few sections.

In giving my alternative theory of content-determination, I don't want to commit myself to any strong claims about our cognitive architecture. I don't, for example, want to commit myself the view that to have a belief (or a desire, or whatever) is to bear a particular relation to a token of a Mentalese sentence. To avoid this commitment, I will take an interpretation function to be a function from propositional attitude-states (belief states, desire states and the rest) to propositions. I remain neutral about what these states are.

2.9 The standard story

There's a widely accepted story about why truth is precious, so widely accepted in fact that I think it deserves to be called 'the standard story'. In this section, I outline and evaluate it.[31]

An example will help. Suppose that Cornelius was feeling unpleasantly tired, and he wanted some coffee. In front of him was a vending machine, on which there was

---

[30] Pg. 125.

[31] My presentation is loosely based on that in Goldman (1999).

a red button marked 'Dispense Coffee'. He came to believe that there was a vending machine in front of him, on which there was a red button marked 'Dispense Coffee'. Cornelius inferred that if he pressed the red button, he would get coffee. So he pressed the button. To his delight, coffee was dispensed.

There are two beliefs to talk about here:

(a) Cornelius's belief that if he pressed the red button he would get coffee.

(b) Cornelius's belief that the red button was marked 'Dispense Coffee'.

It seems to be pretty clear why belief (a) was valuable to Cornelius. He wanted coffee, and he believed that he would get it if he pressed the red button. So he pressed the red button. Since his belief was true, he got what he wanted.

Let's use the term 'means-ends propositions' for propositions of the form 'If I do X, E will happen'. It seems to be pretty clear why it is often useful to believe a true means-ends proposition. Believing a true proposition of the form 'If I do X, E will happen' can allow you to achieve some end E by doing the relevant action X.

Now to belief (b). This belief was useful to Cornelius in a slightly less direct way. It was useful to him because he inferred (a) from it, and (a) was useful to him for reasons already discussed. The point generalises. True beliefs are useful because we can infer true means-ends propositions from them.

This is the standard story. Believing true means-ends propositions is useful because it leads one to take appropriate actions to achieve one's goals. Believing other true propositions is useful indirectly because true means-ends propositions can be inferred from them. All of this sounds very plausible; but is our problem solved? One cause for suspicion is that nothing has so far been said about the crucial question, 'What makes the standard interpretation function so special?'

To get a better hold on what is going on here, I would like to draw attention to three important assumptions implicit in the standard story. First, it is assumed that

an agent who believes 'If I do X, then I will achieve end E' and desires E sufficiently strongly will respond by doing X. More precisely:

> (1) If an agent believes a proposition of the form 'If I adopt means M, then I will achieve end E,' desires E more than anything else in the picture and is in a position to adopt means M, she will adopt means M.[32]

Second, the proponent of the standard story thinks it good to believe true propositions not of the means-ends form, because one will tend to derive true means-ends propositions from them. This clearly assumes:

> (2) Agents usually make reliable inferences: i.e. if an agent has one belief-state, and forms another belief-state on the basis of an inference from the first, and the first belief-state is true, then usually the second is true too.

The final assumption is rather less obvious. The proponent of the standard story thinks that it is useful to believe true means-ends propositions because we are then likely to act in such a way that we get what we want. This clearly assumes:

> (3) An agent usually desires things which are good (or good for her).[33]

I pick out these assumptions for discussion not because I think they are false, but because, if true, they mark peculiarities of the standard interpretation function. To see this, consider the following claims, which parallel (1)-(3):

> (1*) If an agent BELIEVES* a proposition of the form 'If I adopt means M, then I will achieve end E,' DESIRES* E more than anything else in the picture and is in a position to adopt means M, she will adopt means M.

[32]The wording comes from Goldman (1999).

[33]The difference between the two versions is not relevant here.

(2*) If an agent has one belief-state, and forms another belief-state on the basis of an inference from the first, and the first belief-state is TRUE*, then usually the second is TRUE* too.

(3*) An agent usually DESIRES* things which are good (or good for her).

All three of these claims are likely to be false. I can illustrate this by looking at the example again. Suppose, just to keep the example simple, that I* is like the standard interpretation function except that:

Beliefs/desires etc. which concern coffee according to the standard interpretation function, concern pond water according to I*.

Beliefs/desires etc. which concern red according to the standard interpretation function, concern blue according to I*.

In the story, Cornelius BELIEVED* that if he pressed the blue button he would get pond water, and he DESIRED* pond water, so he pressed the red button—this is not the behaviour one would expect if (1*) were true. Also, Cornelius started off BELIEVING* that:

The blue button is marked 'Dispense Coffee'.

from which he inferred a (*de se*) BELIEF*:

If I press the blue button, I will get pond water.

So his first belief-state is TRUE*, but the second is not. The inference is not, as one might put it, 'RELIABLE*'. Finally, Cornelius DESIRED* pond water, which was not good for him, contrary to (3*).

This is, of course, only an example. What is more, it's an example cooked up specifically to make the point. However, I hope that it illustrates that while the assumptions (1), (2) and (3) needed to make the standard story work are plausible, the corresponding claims (1*), (2*) and (3*) are probably false.

So in giving the standard story, one succeeds in explaining why an agent should believe truths, but one makes certain assumptions—(1), (2) and (3)—which themselves cry out for explanation. Why should (1), (2) and (3) hold good, but not (1*), (2*) and (3*)? Assumption (3) is particularly questionable in this context, since as we saw when discussing Goldman's response to Stich, an argument similar to Stich's seems to show that there is nothing normatively important about desire-satisfaction.

2.10 The appeal to decision theory

Several philosophers have attempted to clarify the claim that truth is valuable, and demonstrate its correctness, using decision theory.[34] I will present one such attempt. My discussion will not turn on the details of the decision-theoretic result that I will outline, and so I take it that my discussion would apply equally well to other decision-theoretic approaches to the issue.

I will compare two stories. The protagonist of the two stories is a Bayesian agent with credence function C. In the first story, the agent is presented with a decision problem: she will be asked to choose between options $d_1, \ldots, d_n$, on the assumption that her utility upon choosing d_i if H_j is true will be u_{ij}, where $\{H_1, \ldots, H_m\}$ is a partition. Suppose that, in the first story, the agent chooses option d_X, and let U_1 be her expected utility on the assumption that she so chooses. Suppose that

[34]See in particular Loewer (1993), in which Loewer claims to refute Stich using decision theory.

$\{E_1, \ldots, E_k\}$ is another partition. Then:

$$\begin{aligned} U_1 &= \sum_{\substack{r=1,\ldots,k \\ j=1,\ldots,m}} C(E_r \wedge H_j) u_{Xj} \\ &= \sum_{\substack{r=1,\ldots,k \\ j=1,\ldots,m}} C(E_r) C(H_j \mid E_r) u_{Xj} \\ &= \sum_{r=1,\ldots,k} C(E_r) \left(\sum_{j=1,\ldots m} C(H_j \mid E_r) u_{Xj} \right) \end{aligned}$$

So much for the first story. The second story is like the first, but with an additional complication. In the second story, the agent is to be told, just before she is given the decision problem, which element of $\{E_1, \ldots E_k\}$ is true. Let U_2 be the agent's expected utility in the second version of the story.

Let's calculate U_2. Suppose that, in the second story, the agent is informed that it is E_r which is true. Then she will choose whichever d_i maximises the following quantity:

$$\sum_{j=1,\ldots,m} C(H_j \mid E_r) u_{ij}$$

So her expected utility, on the assumption that E_r is true, is:

$$max_{i \in \{1,\ldots,n\}} \left(\sum_{j=1,\ldots,m} C(H_j \mid E_r) u_{ij} \right)$$

So:

$$U_2 = \sum_{r=1,\ldots,k} C(E_r) max_{i \in \{1,\ldots,n\}} \left(\sum_{j=1,\ldots,m} C(H_j \mid E_r) u_{ij} \right)$$

Say that a proposition E_r is 'inapposite' for the agent if either $C(E_r) = 0$ or d_X would still be a utility-maximising decision after updating E_r. From the expressions for U_1 and U_2 above, we can see that $U_1 \leq U_2$, with equality only if each E_r is inapposite.

In rough summary then, when a rational agent faces a decision problem, her expected utility rises given the prospect of learning an apposite truth.

This looks like a sharpened up version of the claim that truth is valuable—and it has been demonstrated using standard decision theory. Have we shown that there

is something normatively special about truth, after all? Our suspicions should be aroused by the fact that these equations say nothing at all about what makes the standard interpretation function so special.

Just as different interpretation functions can be used to define various belief-like relations, so they can be used to define different credence-like relations. For example, the state of having credence 0.7 in the proposition that badgers eat lettuce might also be the state of having CREDENCE* 0.7 that badgers eat cabbage, and it might also be the state of having CREDENCE** 0.7 that Innsbruck is in Austria. In this way, each person has a CREDENCE* function as well as a credence function. In the same way, each person will have a UTILITY* function and a UTILITY** function, along with her utility function. Stich would of course challenge us with the question, 'why think that credence is of any more interest than CREDENCE*?' It is puzzling how the equations above could help with this question, given that the very same calculations could be performed using the CREDENCE* and UTILITY* functions.

To see what's going on here, we need to notice that in giving the decision-theoretic account of why conditionalising on a true proposition is valuable, we made some assumptions. In particular:

(i) The agent's credence function is 'coherent', in the sense that it is a probability function.

(ii) The agent acts so as to maximise expected utility, where the expected utilities are calculated using her credences.

(iii) It is good (for the agent?) when her utility is maximised—i.e. the agent attaches higher utilities to outcomes that are genuinely good.

I highlight these assumptions not because I think they are false, but because if true they tell us something important about the standard interpretation function.

If we were to try to make a parallel case for the value of CONDITIONALISING* on a truth, we would have to invoke the parallel assumptions:

(i*) The agent's CREDENCE* function is 'coherent', in the sense that it is a probability function.

(ii*) The agent acts so as to maximise expected UTILITY*, where the expected UTILITIES* are calculated using her CREDENCES*.

(iii*) It is good (for the agent?) when her UTILITY* is maximised—i.e. the agent attaches higher UTILITIES* to outcomes that are genuinely good.

Whether or not the agent's credences are coherent may be independent of interpretation, so (i) and (i*) may be equivalent. However, consideration of the coffee/pondwater example suggests that (ii*) and (iii*) are likely to be false. (ii) and (iii) are substantial assumptions, and require explanation themselves.

I like to think of this decision theoretic approach to the issue as a smartened up version of the standard story. The similarities between the two approaches are obvious: the standard story purports to show that having true beliefs helps us get what we want, while the decision theoretic approach purports to show that conditionalising on a truth improves expected utility. Both explanations make use of assumptions which themselves require explanation.

2.11 Rationality and the standard interpretation function

Think about the standard story again. This is an account of why it is so often valuable to believe truths. I argued in section nine that it makes use of certain assumptions, which themselves require explanation:

> (1) If an agent believes a proposition of the form 'If I adopt means M, then I will achieve end E,' desires E more than anything else in the picture

and is in a position to adopt means M, she will adopt means M.

(2) Agents usually make reliable inferences: i.e. if an agent has one belief-state, and forms another belief-state on the basis of an inference from the first, and the first belief-state is true, then usually the second is true too.

(3) An agent usually desires things which are good (or good for her).

These three assumptions have something in common: they are all rationality assumptions.

I think it will not be controversial that rational agents typically behave in the way described by (1), so if people are usually rational, (1) will be true in consequence. Similarly, it seems that rational agents typically make good inferences, from an epistemic point of view. I take it that such inferences are typically truth preserving, and so if most agents are rational (2) will be true in consequence.

(3) requires a bit more discussion. I think it fairly uncontroversial that rational agents will typically desire things which are good (or good for them). Some will derive this from 'subjective' theories of goodness—according to which what's good for a person just is to get what she desires, or what she would desire if fully informed, or something similar. Others will regard (3) as a more substantial claim, about the alignment of a rational person's desires with objective facts about goodness. I will remain neutral about this. All that is important for present purposes is that rational agents tend to desire what is good (or good for them). There are, I think, some exceptional cases. If a rich man offers me a large sum of money on the sole condition that I come to desire there to be a hurricane on the first of August of the year 3000, it will be rational for me to form this desire. So there could be cases in which it is rational for me to desire something which is not good or good for me. Such cases, however, are extremely rare. Typically, rational agents desire the good.

We are used to making absolute judgements of rationality. However, it also makes sense to make judgements of the rationality of an action relative to this or that interpretation function. For example, Cornelius's pushing the red button (in the story) is rational relative to the standard interpretation function (since he believed that so doing would get him coffee, and desired coffee) but was not rational relative to the alternative interpretation function. An action is rational outright just in case it is rational relative to the standard interpretation function. (1), (2) and (3) are true, I think, because people are usually rational relative to the standard interpretation function. (1*), (2*) and (3*) would be true if people typically acted rationally relative to the relevant alternative interpretation function—but they don't.

A similar point can be made by considering the decision-theoretic version of the standard story. The assumptions made in this case are:

> (i) The agents credence function is coherent, in the sense that it is a probability function.
>
> (ii) The agent acts so as to maximise expected utility, where the expected utilities are calculated using her credences.
>
> (iii) It is good (for the agent?) when her utility is maximised—i.e. the agent attaches higher utilities to outcomes that are genuinely good.

I suggest that these are true because people are rational relative to the standard interpretation function. Since people are not usually rational relative to alternative interpretations functions, (i*), (ii*), (iii*) and the like are not generally true. Man is a rational animal, but not often also RATIONAL*.

Here's the upshot of this section. The standard-story provide an account of why it is good for an agent to have true beliefs provided that the agent is otherwise rational, relative to the standard interpretation function. To provide a complete account of why it is good to believe truths, it suffices to explain why people tend to

be rational relative to the standard interpretation function (and not relative to other interpretation functions). I will sketch such an explanation in the remainder of the chapter.

2.12 A way forward

Most of the time, we identify belief-states using that-clauses: we identify them via their (standard) interpretations. There are exceptional cases, of course—one sometimes says things like 'The belief which John expressed with that last sentence', or 'The belief which brought John to tears'—but such cases are rare. The standard interpretations of a person's belief-states, then, are like labels, used to identify them. Alternative interpretation functions are like alternative labelling systems. Looking at it this way reveals an important disanalogy between the standard interpretation function and the alternatives: the alternative interpretation functions are for the most part terrible as labeling systems. Let's look at the coffee example again. Here are the relevant causal interactions between mental-states, described using their standard labels:

> (a) Cornelius was unpleasantly tired, so he desired coffee.
>
> (b) Cornelius was in the presence of a red button labeled 'Dispense Coffee', and this caused him to believe that he was in the presence of a red button labelled 'Dispense Coffee'.
>
> (c) Cornelius's belief that he was in the presence of a red button labeled 'Dispense Coffee', caused him to believe that if he pressed the red button, he would get coffee.
>
> (d) Cornelius's belief that if he pressed the red button he would get coffee, together with his desire for coffee, caused him to press the red button.

Now let's describe these same causal interactions again, using the alternative labeling system:

> (a*) Cornelius was unpleasantly tired, so he DESIRED* pond-water.
>
> (b*) Cornelius was in the presence of a red button labeled 'Dispense Coffee', and this caused him to BELIEVE* that he was in the presence of a blue button labeled 'Dispense Coffee'.
>
> (c*) Cornelius's BELIEF* that he was in the presence of a blue button labeled 'Dispense Coffee', caused him to BELIEVE* that if he pressed the red button, he would get pond-water.
>
> (d*) Cornelius's BELIEF* that if he pressed the blue button he would get pond-water, together with his DESIRE* for pond-water, caused him to press the red button.

Described this way, these processes seem completely haywire. Why should tiredness cause Cornelius to DESIRE* pond-water? And why should the presence of a red button labeled 'Dispense Coffee' cause him to come to BELIEVE* that if he pressed a blue button he would get pond-water? This is a bad labeling system.

We should look at this in more detail: what is it that makes the standard labeling system so much better than the alternative?

The 'causal role' of a mental state, as I use the term, is characterised by statements which describe what does or would cause the state, and what it does or would cause. For example, the causal role of the belief that one is tired is described by saying that it tends to be caused by tiredness, and it tends to cause people to go to bed, or look for stimulants. The non-standard labels for the mental states in the story are bad, because they have nothing to do with the causal roles of those states. In consequence, they make the task of predicting, explaining and evaluating a person's actions completely unmanageable. By contrast, the causal roles of these same mental

states are readily predictable on the basis of their standard labels. Why? Because relative to the standard interpretation function people act rationally. Relative to the alternative interpretation function, they don't. What I'm suggesting is that, contrary to Stich, our intuitions about truth are not just 'an idiosyncratic hodgepodge bequeathed to us by our cultural and/or biological heritage.' On the contrary, we label belief-states in a systematic way so that the causal role of a given belief state is predictable on the basis of its label, its interpretation.

I should set out my proposal more explicitly. Here's a preliminary point: rationality comes in degrees. It makes sense not only to assess an action as rational or irrational; one can also say that one action was more rational (or even much more rational) than another. When Cornelius (still on a diet) reaches the end of his supper, the most rational thing for him to do is eat an apple. A slightly less rational thing for him to do is eat a sticky bun. Still less rational would be to drink washing up liquid. So here's the proposal. One can describe an agent's behavioural dispositions using counterfactuals such as this one:

> If a horse were to step on Cornelius's foot, he would say 'Blast!'.

The actions described in the consequents of these counterfactuals are actions the agent is disposed to perform, as I like to put it. The standard interpretation function, I suggest, is whichever interpretation function maximises the average degree of rationality of the actions the agent is disposed to perform. To put it slightly differently, the actions the agent is disposed to perform are on average more rational relative to the standard interpretation function than relative to any alternative interpretation function.[35]

[35]If there is a tie for first place, the result is vagueness or open texture.

Of course, this proposal is not original to me. It's a variant on David Lewis's theory of content-determination,[36] and similar views have been defended by many other philosophers. It is controversial whether such views are ultimately defensible, when the details are worked out—and this is not the place to contribute to the debate. But I hope to have convinced you that it is an advantage of the rationality-maximising theory of content determination that it allows us to explain why it is good to believe things which are true.

[36]See in particular Lewis (1974).

Chapter 3

Analyticity in Mathematics, Part One

3.1 Introduction

Loosely speaking, the operator '$Nx : \ldots x \ldots$' means 'the number of things x such that $\ldots x \ldots$'. For example, '$Nx : dog(x)$' refers to the number of dogs, while '$Nx : (cat(x) \wedge black(x))$' refers to the number of black cats. Neofregeans claim that this operator can be defined implicitly with the following formula, 'Hume's Principle':[1]

(HP) $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$

Here, '$F \sim G$' abbreviates a formula which means 'there is a one-to-one correspondence between the Fs and the Gs', or more simply, 'there are exactly as many Fs as Gs'.[2] So (HP) as a whole expresses the claim that, for any F and G, the number of

[1]In this paper, I'm discussing a view on which a written symbol 'N' is defined by a sentence. There's a related view on which what's defined is a concept rather than a symbol, and the definition is a belief rather than a sentence. Much of what I say about the 'linguistic' version of neofregeanism also applies to the 'mentalistic' version, though to keep things simple I'm only discussing the former.

[2]Specifically, $\exists R[\forall x(Fx \rightarrow \exists y(Gy \wedge \forall z(Rxz \leftrightarrow z = y))) \wedge \forall y(Gy \rightarrow \exists x(Fx \wedge \forall z(Rzy \leftrightarrow z = x)))]$.

Fs equals the number of Gs just in case there are exactly as many Fs as Gs. Using the 'N' operator, other numerical terms can be defined explicitly. For example, we can define 'zero' or '0' as the number of things not identical with themselves:

$$0 =_{def} Nx : x \neq x$$

And we can define 'number' by saying that something is a number just in case, for some F, it is the number of Fs:

$$\forall y[number(y) \leftrightarrow \exists F(y = Nx : Fx)]$$

It is also possible to define 'successor' and 'natural number' explicitly using the 'N' operator.[3]

This seems to provide the beginnings of an attractive explanation of our competence with numerical concepts. However, for many people the big draw of the neofregean claim that (HP) is an implicit definition of the 'N' operator is epistemological.

In his book *Frege's Conception of Numbers as Objects*,[4] Crispin Wright proved that the standard axioms of number theory, the Peano axioms, are derivable from (HP) given the definitions of '0', 'natural number' and 'successor' mentioned above. Wright suggested that because (HP) is a definition, it is knowable *a priori*. Moreover, he claimed, since the Peano axioms are provable from (HP), they are *a priori* too, and so the theorems of Peano arithmetic are also *a priori*. If this is correct, Wright has explained the apriority of number theory.[5] Wright has subsequently defended

---

[3]Here's a definition of 'Successor':

$$Successor(n_1, n_2) \leftrightarrow \exists F[n_1 = Nx : Fx \wedge \exists y[\neg Fy \wedge n_2 = Nx : (Fx \vee x = y)]]$$

One can define 'natural number' by stipulating that the natural numbers are those things which to which zero bears the ancestral of the successor relation.

[4]Wright (1983).

[5]This assumes that a statement that is entailed (in second order logic) by *a priori* truths is itself *a priori*. That's an assumption that deserves scrutiny: but I won't address the issue here. Neofregeans

this position in a series of papers, many of which were written in collaboration with Bob Hale.

In defending this position, Wright and Hale oppose Kant's widely accepted claim that one cannot obtain *a priori* knowledge of an existential generalisation by deducing it from definitions.[6] If Kant was right about this, neofregeanism is a non-starter, for (HP) implies that all of the following exist and are distinct:

$$0 := Nx : x \neq x$$

$$1 := Nx : x = 0$$

$$2 := Nx : (x = 0 \vee x = 1)$$

$$3 := Nx : (x = 0 \vee x = 1 \vee x = 2)$$

$$\ldots$$

$$\ldots$$

It is striking how widely intuitions differ on this issue. Some people find it highly plausible that (HP) is an implicit definition. Other people think it completely obvious that no definition may have existential import. One of my goals in this paper is to 'get behind' these intuitions, and to explain their divergence. I will suggest that philosophers on the two sides are working with different accounts of how definitions work. The Kantians assume what I call 'the reference based' theory, while the neofregeans rely on the 'truth based' theory. My second goal is to develop, on behalf of the neofregeans, a detailed version of the truth based theory, which they

---

break down the problem of explaining the apriority of mathematics into two sub-problems. The first sub-problem is explaining the apriority of 'basic' mathematical truths. They attempt to solve this problem by saying that the basic mathematical truths are definitional. The second sub-problem is that of explaining why 'entailment preserves apriority'. I won't be looking at this second sub-problem here. That's an issue for another day.

[6]See book II chapter III section 4 of Kant (1781).

can employ against the 'Kantian' objection that (HP) can't be a definition because it entails existential generalisations.

3.2 The reference based theory

Let's suppose that we introduce the name 'Goliath' with this definition:

> Goliath is a man at least 10cm taller than every other man.

It seems that one could only make this sentence true by stipulation on the condition that some man is at least 10cm taller than every other man. If there is no such man, the definition will 'fail': the sentence won't become true. As I will put it, the 'success condition' for this definition is that some man is at least 10cm taller than every other man. More generally, the success condition of a definition is the condition that needs to be met in order for the definition to succeed.

Here's a scientific example, the standard SI definition of 'metre':

> A metre is the distance that light travels in a vacuum in one 299,792,458$^{\text{th}}$ of a second.

Before adopting this definition, physicists had to be sure that the speed of light in a vacuum is constant. If the speed of light had not been constant, they couldn't have made this statement true by stipulation. The definition would have failed. The success condition of the definition is that the speed of light in a vacuum is constant.

Just to be clear on the terminology, when a definition fails, it's still a definition. I assume that, provided one has appropriate linguistic authority, one can always make a sentence into a definition by stipulation. However, one can only make a sentence true in this way if the definition's success condition is met. I suppose that one could use the word 'definition' in a more restrictive way, so that there are no failed definitions. But that's not my preferred way of doing it.

The reference based theory is an account of what the success conditions of a definition are. The basic idea is simple enough. According to the reference based theory, the purpose of a definition is to assign a referent to the newly introduced term or terms; the definition will succeed just in case a suitable referent exists, or just in case suitable referents exist. Different versions of the reference based theory may differ to an extent on what they imply about what 'suitability' amounts to, but the general idea is that when someone defines a new term, she is taking direct conscious control over its meaning. She forms certain intentions about how the term is to be understood; a suitable referent for the new term would be one which accords with those intentions. In the sort of cases that interest us, the agent's intention will be to make a certain sentence, the definition, express a truth, or sometimes a necessary truth. She will also typically want all other terms in her language to retain their normal referents.

In short, then, the reference based theory is that a definition succeeds just when referents exist for the newly defined terms that are 'suitable', in the sense that they accord with the agent's intentions.[7] In the 'Goliath' example, the intention of the agent is to ensure the truth of this sentence:[8]

> Goliath is a man at least 10cm taller than every other man.

In consequence, a suitable referent for 'Goliath' would be a man at least 10cm taller than every other man. So, according to the reference based view, the success condition of the definition is that there exists such a man. The reference based theory works rather well in this case. It also does well in the case of 'metre'.

---

[7] I suppose that when *several* suitable referents exist, the referent of the newly defined term will be correspondingly indeterminate; however, to avoid getting in to the theory of vagueness I will say little about this issue.

[8] To avoid unnecessary complexity, I'm avoiding ignoring issues to do with context-sensitivity here. I'm not paying attention to the fact that the definition is in the present tense, and so is context-sensitive. Nothing important changes when we take context-sensitivity into account.

Now let's look at the implications of the reference based theory for neofregeanism. Here's (HP) again:

(HP) $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$

In using (HP) as a definition of the 'N' operator, one's intention is presumably that (HP) be necessarily true. Now it might be thought that the defined term here is the 'N' operator, and so the success condition for (HP) is that there exists a suitable function from concepts to objects. But this is a bit quick. The neofregean could consistently deny that the 'N' operator refers. However, the neofregean is committed to the claim that terms of the form $\ulcorner Nx : \phi(x) \urcorner$ refer, because she wants to licence the inference:

$\ldots Nx : \phi(x) \ldots$; therefore $\exists y \ldots y \ldots$

Inferences like this are crucial if a substantial body of mathematics is to be derived from (HP). In consequence, assuming the reference based theory, (HP) will succeed only if suitable referents exist for expressions of the form $\ulcorner Nx : \phi(x) \urcorner$. Now it is not completely obvious what some objects would have to be like in order to be suitable referents for such terms (Would they have to be abstract? Would they have to exist necessarily?), but it is clear that there would have to be infinitely many of them, for as I said (HP) implies that 0, 1, 2, 3 and so on exist and are distinct.

So while it is not clear exactly what implications the reference based theory has about the success conditions of (HP), the reference based theory does imply at least that (HP) will succeed only if there exist infinitely many objects.

3.3 The infinity problem

Now let's talk epistemology. Suppose one introduces a new term by stipulating that a certain sentence containing the term is to be true. It seems then that one will know

the sentence to be true only if one knows independently that the success condition of the definition is met. Consider for example the definition of 'Goliath':

> Goliath is a man at least 10cm taller than every other man.

Having introduced this definition, one will know that it is true only if one knew in the first place that some man is at least 10cm taller than every other man. To see what I mean by 'independently' and 'in the first place', imagine meeting someone who claims to know that the definition of 'Goliath' is true. You ask if she knows that the success condition of the definition is met, and she says that she does. You ask her *how* she knows this, and she explains that the success condition of the definition is that some man is at least 10cm taller than every other man, and she knows that this condition obtains because it's an analytic truth—it follows from the definition of 'Goliath'. This is surely ridiculous: one cannot know *a priori* that the world's tallest man is at least 10cm taller than the closest runner up. To know that the definition is true, one must know *independently of the definition* that some man is taller by 10cm than every other man.

I'll have more to say about this in section eight, but for now my provisional conclusion is that one can know a definition to be true only when one knows independently that its success condition is met. I'll go further: when one does know that a definition is true, one knows this in part *because* one knows that its success condition is met. One's knowledge of the truth of the success condition is part of one's justification for believing the definition itself. If this is right, then one can know *a priori* that a definition is true only if one can know independently and *a priori* that its success condition obtains.

This has important implications for the epistemology of neofregeanism. It implies that one can know *a priori* that (HP) is true only if one knows independently and *a priori* that its success condition obtains. Assuming the reference based theory, this means that (HP) is *a priori* only if one could know *a priori* and independently of

(HP) that there exist infinitely many things. And it is not clear how one could know this.

Here's Ted Sider, summarising the problem:

> How can something that implies the existence of even one thing, let alone infinitely many, be a definition? What the neofregeans claim is that, despite its existential implications, Hume's Principle is nothing more than a definition of number, and therefore can be known to be true *a priori*. Thus, in a sense, objects may be introduced by definition. ... Think of the act of laying down the definition as the delivery of instructions to the semantic gods: let my expression 'the number of' be so understood as to obey Hume's principle. But of course, this just invites the question of whether there is any way to understand 'the number of' so that Hume's Principle comes out true. If there do exist infinitely many objects, then perhaps there is a way, but if there are not, one wants to say, there may simply be no way of interpreting 'the number of' so that Hume's Principle comes out true. In that case, the semantic gods will respond to our instructions with a blank look, as they would (assuming atheism) if we stipulated that 'God' is to denote the omnipotent being who created the world.[9]

It might be replied that we can know *a priori* and independently of (HP) that there exist infinitely many things, because this is an implication of set-theory, which is an *a priori* discipline. However, Hale and Wright would not be happy with this approach: their goal is to provide an epistemological basis for arithmetic independent of other parts of mathematics. As I will put it, according to the Hale and Wright position, (HP) is *mathematically basic*—it can be known without being derived from other mathematical claims.

I will now present the argument step-by-step. The first premise is:

> (1) (HP)'s success condition is at least that there exist infinitely many things. (From the reference based theory of definition)

(HP) is supposed to determine referents for all expressions of the form $\ulcorner Nx : \phi(x) \urcorner$, and in particular to '0', '1', '2' and so on, as defined above. (HP) implies that all of

[9]Sider (2007).

these objects are distinct, and so (HP) will be true only on the condition that there are infinitely many things.

Then:

> (2) (HP) can be mathematically basic and knowable *a priori* only if it is knowable *a priori* and independently of (HP) and other mathematical claims that (HP)'s success condition is met. (Premise)

Hence:

> (3) (HP) can be mathematically basic and knowable *a priori* only if it is knowable *a priori* and independently of (HP) and other mathematical claims that there exist infinitely many things. (From (1), (2))

Finally:

> (4) It cannot be known *a priori* and independently of (HP) and other mathematical claims that there exist infinitely many things. (Premise)
>
> (5) (HP) cannot be mathematically basic and known *a priori*. (From (3), (4))

I call this 'the infinity problem' for the neofregeans.

I should pause to rebut an objection to the argument. Recall that in making the case for (1) I pointed out that (HP) implies that the following are all distinct:

$$0 := Nx : x \neq x$$

$$1 := Nx : x = 0$$

$$2 := Nx : (x = 0 \vee x = 1)$$

$$3 := Nx : (x = 0 \vee x = 1 \vee x = 2)$$

$$\ldots$$

$$\ldots$$

This assumes that there are infinitely many type expressions containing the 'N' operator—and hence infinitely many things. This may seem objectionable: isn't the whole point to call into question this assumption? My response is that in making this argument against the neofregeans, one need not deny that there are infinitely many things. What is at issue is not whether there exist infinitely many things, but whether this is knowable *a priori* because it is deducible from definitional truths.

3.4 Introducing the truth based theory

As far as I know, Hale and Wright never explicitly reject the reference based theory of definition. However, when describing their own position they invariably describe (HP) as determining the truth conditions of sentences containing the 'N' operator, rather than as assigning referents to singular terms of the form $\ulcorner Nx : \phi(x) \urcorner$. I suggest, then, that we understand Hale and Wright as rejecting the reference based theory of definition. According to their position, in the first instance at any rate, (HP) fixes the truth conditions of sentences rather than the referents of sub-sentential expressions. It ensures, for example, that the second of these two sentences has the same truth condition as the first (which of course already has a truth condition, independent of the stipulation):

> There are exactly as many cities in Wales as species of rhinoceros.
>
> The number of cities in Wales = the number of species of rhinoceros.

Now as it happens, there are indeed exactly as many cities in Wales as there are species of rhinoceros, so the stipulation ensures that the second sentence is in fact true. By existential generalisation, we can infer that there does indeed exist a number (viz. the number of cities in Wales, which is 5).

If we assume the relevant disquotational sentence:

'The number of cities in Wales' refers to 5 $\leftrightarrow$ The number of cities in Wales=5.

we can infer that the stipulation does in the end ensure that 'the number of cities in Wales' refers, but this is merely a consequence of the more basic effect of (HP), which is to determine the truth conditions of statements containing the 'N' operator. Truth conditions first; reference second. Hale and Wright, if I understand them properly, would like to reject the reference based theory of definition in favour of a truth based theory: a theory according to which a definition succeeds just in case there exists a suitable assignment of truth conditions to sentences containing the defined term. They can reject the argument presented in the last section at its first premise.[10]

Later in the paper, I'm going to develop this basic idea into something more solid. In the rest of this section, I am going to discuss some *prima facie* obstacles to the response.

One problem was raised by Frege himself in response to the suggestion that (HP) is a definition.[11] Frege pointed out that (HP) doesn't seem to determine the truth conditions of all sentences containing the 'N' operator, because it doesn't determine, for example, the truth conditions of:

$(Nx : x \neq x) =$ Julius Caesar.

More generally, (HP) doesn't seem to determine the truth conditions of identity statements with an 'N' term on one side of the identity sign and a singular term of

[10]Hale and Wright have discussed the 'Kantian' idea that existential generalizations can't be true by definition in a number of places, and I confess that I don't find everything they say very easy to follow. See Hale and Wright (2000) and Hale and Wright (2009b). See also Wright's discussion of Frege's Context Principle in Wright (1983). Sider (2007) contains some critical discussion.

[11]See Frege (1884). The relevant passage is on pg. 68 of Austin and Frege (1974).

another kind on the other. This is known as the 'Caesar problem'. It may seem to be an odd complaint: these statements aren't very important from a mathematical point of view, so who cares? However, Frege's goals weren't only mathematical. He wanted a theory which would explain *what numbers are*, and so also, what they aren't.

Here in outline is Hale and Wright's response. Say that a 'sortal' predicate is a term associated with a 'criterion of identity' for the things of which it is true. A criterion of identity for some objects is a specification of necessary and sufficient conditions for identity statements concerning those objects, conditions which explain what identity and non-identity 'consist in' for such objects. For example, the concept set is a sortal, and the associated criterion of identity is that two sets are identical just in case they have the same elements.

Hale and Wright say that we should not introduce the 'N' operator using (HP) alone. Rather, we should stipulate something slightly stronger, something more like:

> (HP^2) 'Number' is a sortal predicate, and the corresponding criterion of identity is: $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$

Hale and Wright then appeal to a general metaphysical principle, the 'sortal exclusion principle', according to which nothing can fall under two sortal predicates. We can tell, they argue, that Julius Caesar is not a number by observing that Julius Caesar falls under the sortal predicate 'person'[12] and then appealing to the sortal exclusion principle to establish that he does not also fall under the term 'number'.[13]

Closely related to the Caesar problem is what Kit Fine calls the 'Roman Problem'.[14]

---

[12]...or 'organism' or 'physical object' or something else—not 'number', at any rate.

[13]The term 'sortal' is ambiguous, and the sortal exclusion principle will not plausible on every reading of the term.

[14]See Fine (2002).

(HP) doesn't seem to determine the truth conditions of statements like:

$(Nx : x \neq x)$ is a Roman emperor.

More generally, (HP) doesn't determine the truth conditions of statements that contain a non-mathematical predicate applied to an 'N' term.

Presumably Hale and Wright would appeal to the sortal exclusion principle again in response. It's obvious, they would say, that all Roman emperors are people. By the sortal exclusion principle, it follows that numbers are not Roman emperors, so in particular $(Nx : x \neq x)$ is not a Roman emperor.

So Hale and Wright have responses to the Caesar and Roman problems. But there are further problems. Hale and Wright say that the stipulation of (HP) ensures that these two statements have the same truth-condition:

There are exactly as many cities in Wales as species of rhinoceros.

The number of cities in Wales = the number of species of rhinoceros.

Fine. But what about instances of (HP) which have the 'N' operator on both sides of the conditional? For example:

$$(Nz : (\lambda x.x \neq x)z) = (Nz : (\lambda x.x = (Ny : y \neq y))z)$$
$$\leftrightarrow \qquad (\lambda x.x \neq x) \sim (\lambda x.x = (Ny : y \neq y))$$

It is not enough to say that (HP) ensures that the left-hand formula has the same truth condition as the right-hand formula, because the right-hand formula has no truth condition independent of (HP). These instances of (HP) are crucial to the derivation of the Peano axioms from (HP), so they cannot be ignored.[15]

[15]Wright discusses the impredicativity of (HP) in Wright (2001a), though in a rather different context. See Fine (2002) for some criticism.

The most serious problem is this. Neofregeans would like to respond to the argument in the previous section by rejecting its first premise, viz., that the success condition of (HP) is at least that there are infinitely many things. However, so far we have not derived from the truth based theory any specific conclusion about what the success condition of (HP) is—and I cannot see how to get such a conclusion without developing the truth based theory in a little more detail. I'll do this over the next three sections.

3.5 Introducing my version of the truth based theory

My proposal is that a definition succeeds just in case there exists an assignment of truth conditions to the sentences in the agent's newly extended language which accords with the agent's intentions. I will identify truth conditions with classes of possible worlds. A function that assigns truth conditions to sentences will be called an 'interpretation',[16] and I'll use the term 'suitable' for interpretations which accord with the agent's intentions. I'm going to assume that there already exists a suitable interpretation function J for the agent's 'old language'—the language she used immediately before the definition. The definition succeeds if it can be replaced by a new suitable interpretation function J^+, which covers sentences containing the newly defined term or terms.

Now so far, we can't draw conclusions about the success conditions of specific definitions from the truth based theory. To do this, we need to spell out in some detail what the agent's intentions are going to be like in the cases that concern us.

[16]It might be objected that because some sentences are ambiguous, J should not be a function. In response, I suggest that (for example) 'John has gone to the bank' is not the same sentence as 'John has gone to the bank'.

I'll do this by spelling out a number of conditions on suitability. The idea is that the suitable interpretation functions for the newly extended language will be precisely those which meet the conditions.

First, it seems that in defining a new term, we typically mean only to give meanings to sentences containing that term. The meanings of 'old' sentences, not containing the new term, should not be affected. This motivates the first condition:

(C1) For any 'old' sentence ϕ, $J^+(\phi) = J(\phi)$.

What about the assignment of truth conditions to the definitions themselves? It is worth distinguishing cases here. In some examples, the agent will intend that the definition be *necessarily true*. The case of (HP) will presumably be like this. In other cases, the agent will intend merely that the sentence be true, but perhaps only contingently so. The 'Goliath' example is in this latter category. I am going to limit my attention to the former sort of case.

This doesn't involve any significant loss of generality, because we can assimilate cases of the latter kind to those of the former kind simply by adding the word 'actually' into the definitions. So for example, we can modify the definition of 'Goliath' as follows:

Actually, Goliath is a man at least 10cm taller than every other man.

The second condition is then straightforward:

(C2) For each of the definitional statements δ, $J^+(\delta)$ is the set of all possible worlds.

Finally:

(C3) J^+ respects entailment, in the sense that if w is an element of $J^+(\gamma)$ for each γ in some set Γ, then w is an element of $J^+(\alpha)$ for any α entailed by Γ.

This condition may be inapplicable in the case of definitions of logical constants. In such cases, perhaps, we have no understanding of entailment independent of the interpretations of sentences containing the defined term. But I am not concerned with these cases here.

I am not going to say very much about what 'entailment' is here, but for the record my position is that the relevant notion of entailment will vary from agent to agent. Some agents may operate with a 'proof-theoretic' notion of entailment. In this case, a statement α will be entailed by a set of formulas Γ just in case there is a proof of α from premises drawn from Γ, where what counts as a 'proof' is defined syntactically. Some other agents may operate with a 'semantic' notion of entailment—that is, one that we can characterise using model theory in the normal way. In the case of agents who use (HP) to introduce mathematical vocabulary into their language for the first time, the agents themselves won't be able to characterise their notion of entailment model-theoretically or proof-theoretically, since models and (in the relevant sense) proofs are both abstract, mathematical objects. However, this doesn't prevent us, from the outside as it were, characterising the relevant notions of entailment in these terms.

I suggest that, for many definitions, these three conditions are enough: satisfying (C1), (C2) and (C3) is necessary and sufficient for suitability. I find these conditions intuitive in their own right: but it is worth stressing that there is a rationale behind all three of them. The idea is that a definition will succeed just in case there is an assignment of truth conditions to the agent's sentences which accords with her intentions.

Let's see what implications this has in the 'Goliath' example. First, notice a consequence of these conditions. Suppose a definition δ entails some 'old' sentence ϕ. If a suitable interpretation function J^+ exists, δ is true relative to J^+ (by (C2)), and δ entails ϕ, so ϕ is true relative to J^+ (by (C3)), and so ϕ is true relative to J (by

(C1)). So if the definition succeeds, any sentence that δ entails that does not contain the new term or terms will be true, as it was understood prior to the definition. Now here again is the (modified) definition of 'Goliath':

> Actually, Goliath is a man at least 10cm taller than every other man.

This implies:

> Actually, some man is at least 10cm taller than every other man.

This sentence doesn't contain the new term. So an interpretation J^+ which meets the three conditions will exist only if this latter sentence is true, as it was understood before the definition was performed. Thus, the success condition of the definition is at least as strong as the claim that some man is at least 10cm taller than every other man. Now if there is such a person, it is fairly straightforward to see that an interpretation meeting the three conditions will exist. So in fact the truth based view implies that the success condition of the definition is precisely that some man is taller than every other man.

So my version of the truth based theory has the same implications as the reference based theory, in the 'Goliath' case. This is as we should hope, since the reference based theory seems to make the right prediction in this case. My version of the truth based theory works well in the case of 'metre' too. Sadly, the theory doesn't work so well in the case of (HP), as we shall see in the next section.

3.6 Creative definitions

Someone might object to these ideas as follows:

> Let's see what happens when we apply these three conditions in the case of (HP). It is easy to construct a sentence (ϕ, let's say) in second-order logic which is true just in case there are infinitely many things; ϕ need

> not contain the 'N' operator. Now (HP) implies ϕ, so your three conditions will be satisfiable only if ϕ is true as it was understood prior to the definition. That is, your view implies that (HP)'s success condition is at least as strong as the proposition that there are infinitely many things. This is problematic from a neofregean point of view. You rejected the reference based theory on the grounds that it has the consequence that (HP)'s success condition entails that there exist infinitely many things; now we find that your truth based theory has this same implication.

I suggest we respond to this by saying that (HP) is a rather different kind of definition to the definitions of 'Goliath' and 'metre': the intentions of the agent are different in the two cases. While (C1), (C2) and (C3) are appropriate for 'Goliath' and 'metre', we need a slightly different set of conditions for the case of (HP). To see what I have in mind, imagine meeting someone who has recently introduced an 'N' operator into her language by stipulating that (HP) is necessarily true, having previously had no vocabulary for talking about abstract objects. Imagine asking her whether the sentence ϕ (which means that there are infinitely many things) was true in her language prior to the stipulation. She might say:

> Until recently, I spoke a 'nominalist' language—I had no way of talking about numbers (or other abstracta). I had no 'N' operator for talking about particular numbers, and my quantifiers ranged only over concrete things. Thus, ϕ was true (in my language then) only if there are infinitely many *concrete* things; I still don't know if this is true.

I said earlier that the motivation for (C1) is that in introducing a new term into one's language, one does not mean to change the meanings of sentences that do not contain the new term. I suggest that the neofregean should say that this is not right in the case of (HP)—though it is correct for the definition of 'metre'. When one

introduces (HP) as a definition, the neofregean should say, one changes the meanings of one's quantifiers, one 'extends' them. To use Kit Fine's phrase, (HP) is a 'creative' definition. This sets it apart from the definition of 'metre'. If this is right, we need to offer a modified version of (C1), to deal with creative definitions.

Here is my suggestion. When one stipulates the necessary truth of (HP), one should also choose a predicate by which to define the domain of the 'old' quantifiers. For any formula ϕ, let ϕ^P be the result of restricting all the quantifiers in using the predicate P. I suggest we modify the first requirement like this:

> (C1*) For any 'old' sentence ϕ, $J^+(\phi^P) = J(\phi)$ (where P is a predicate in the new language identifying the items in the old ontology).

Note that P need not be a predicate in the old language: it may be definable only using the newly defined vocabulary.

It turns out that once we make this modification to (C1), we need to change (C2) as well. It seems clear that any adequate J^+ will assign the set of all possible worlds to 'old' definitional truths (unless of course the agent intends to abandon some old definitions). If, for example, 'Bachelors are unmarried' was definitional prior to the stipulation of (HP), it should still be a necessary truth afterwards. (C1) ensured this, but (C1*) does not (it ensures only that 'All Bachelors that are not numbers are unmarried' is a necessary truth—which is weaker). To avoid this problem, I suggest the following modification to (C2):

> (C2*) For any definitional truth δ (either one of the new definitions, or an 'old' definition), $J^+(\delta)$ is the set of all possible worlds.

I suggest that a definition like (HP), a creative definition, will succeed just in case there exists an interpretation which satisfies (C1*), (C2*) and (C3). It turns out (for a demonstration, see the appendix) that an interpretation meeting these crite-

ria will exist provided that the following condition is met—what I call the 'modal conservativeness' condition:

> Suppose a language L with interpretation function J is extended by way of definitions $\delta_1, \ldots, \delta_m$, and suppose that $\delta_{m+1}, \ldots, \delta_n$ are the agent's existing definitions. Suppose L^+ is the new, extended language. Suppose also that the agent takes it that a predicate P in the newly extended language delimits the 'old' ontology. The modal conservativeness condition is met just in case for any possible world w, $\{\delta_1, \ldots, \delta_n\} \cup \{\alpha^P \in L^+ \colon w \in J(\alpha)\}$ is consistent.[17]

Let's apply this to (HP). Here the relevant predicate P can be defined as follows:

$$\forall x(Px \leftrightarrow \neg\exists F(x = Ny : Fy))$$

So P can be read 'is not a number'. Roughly speaking, the modal conservativeness condition is met provided that for any possible world, (HP) and the agent's old definitions are together consistent with all the truths *about non-numbers* at that world.

All of this should look attractive to the neofregean. If my proposal is correct, the success condition of (HP) is plausibly something that one could know *a priori* and independently of other mathematical claims (more on this in the next section). However, there's a hitch. To see this, consider an English speaker who introduces numbers to her ontology by using (HP) as a definition. The sentence:

> All zebras are stripy.

[17] As I use the term 'consistent', a set of statements is consistent iff it doesn't entail $\bot$.

will presumably be true relative to J, and so relative any adequate J^+. However, assuming that this is not a definitional truth, our conditions are not strong enough to ensure that this is so.

They do ensure that this is true:

> All zebras that are not numbers are stripy.

But this is surely too weak: it seems that our conditions are not strong enough.

This is closely related to the Caesar problem, which I mentioned in the last section. Hale and Wright's respond, you will recall, by saying that (HP) is not quite an adequate definition of the 'N' operator. They suggest a slightly stronger definition:

> (HP^2) 'Number' is a sortal predicate, and the corresponding criterion of identity is: $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$

Hale and Wright then appeal to a general metaphysical principle, the 'sortal exclusion principle', according to which the extensions of sortal terms cannot overlap. We can tell, they argue, that Julius Caesar is not a number by observing that Julius Caesar falls under the sortal term 'person' and then appealing to the sortal exclusion principle to establish that he does not also fall under the term 'number'.[18]

Here's how we might try to apply this to save my version of the truth based theory. We could argue that, by condition (C2*), any adequate J^+ will assign the set of all possible worlds to:

- The sortal exclusion principle.
- (HP^2).
- An appropriate statement of a criterion of identity for zebras.

[18]On this view, a definition may be bad even if it is successful, for it may be that there are *too many* suitable interpretations. In such cases, there will be unwanted indeterminacy in the language after the definition.

These three things together will imply 'No number is a zebra', so by (C3) this will also be assigned the set of all possible worlds. 'All zebras are stripy' is true relative to J, and so 'All zebras that are not numbers are stripy' will be true relative to J^+, by condition (C1*). So by condition (C3) again, 'All zebras are stripy' will be true relative to J^+. Thus, the problem is avoided.

I think that this is basically the right approach to our problem. However, I have my doubts about the sortal exclusion principle—simply because we have rather few good examples of criteria of identity. I don't want to discuss this issue here, but I would like to suggest an alternative to (HP^2) that avoids appeal to identity-criteria.

My suggestion is that we strengthen (HP) by adjoining statements about the metaphysical characteristics of numbers, statements sufficient to distinguish them from items in the 'old' ontology:

(HP^3) $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$

Numbers exist necessarily.

Numbers have no location in space.

Numbers have no location in time.

Numbers are not physical.

...

...

What goes on this list will be influenced on what the agent already has in his ontology. If, for example, the agent's quantifiers already range over sets, 'numbers are not sets' could go on the list.[19] In order to make this work, we will need to suppose (for

[19]It is a consequence of this approach that there will often be a great deal of indeterminacy in mathematical language. If for example the agent introduces both set-theoretical and numerical vocabulary by way of definitions, but fails to specify the truth conditions of identity statements between numbers and sets, such statements will be indeterminate in truth-value. It will also be in many cases indeterminate whether mathematical terms of different languages corefer. Unlike Fine (see Fine (2002), pg. 70) I don't see this as at all problematic. On the contrary, in fact.

example) that it is definitional that zebras are physical. We can then argue, as above, that 'All zebras are stripy' is true relative to any adequate interpretation function for the newly extended language.

Notice that both of the two proposed solutions to the problem involve substantial commitments about definitionality; according to the first proposal, the sortal exclusion principle is definitional; according to the second 'Zebras are physical' (or something similar) is definitional. Many will regard this as a cost of neofregeanism.

3.7 Solving the infinity problem

I've now finished presenting my truth based account of the success conditions of definitions. I have yet to use this to confront directly the argument against neofregeanism presented in section three, and I have not said much about what follows from the truth based account about the epistemology of definition.

I'll start with the latter task. I stand by the idea that in order to know that one's definition is true it is enough that one have some independent reason for thinking that its success condition is met. (See section eight for some further discussion of this claim). If the truth based theory is correct, then, it suffices, in the case of creative definitions, to know that one's definitions are modally conservative. As I said back in section one, I am going to avoid questions about the epistemology of logic in this paper, so I will not have much to say about how one could know that this condition is met. I suppose, however, that this is something one can learn by trying unsuccessfully to derive contradictions from one's definitions together with modally consistent empirical claims,[20] by investigating (HP)'s entailments, trying to get a clear picture of what a model for (HP) would be like,[21] by investigating statements

[20]By 'modally consistent' I mean this. $\alpha_1, \ldots \alpha_n$ are modally consistent just in case $\Diamond(\alpha_1 \wedge \ldots \wedge \alpha_n)$.

[21]I mean 'model' in an informal sense here—I don't mean to say that in order to establish that (HP)

logically equivalent to (HP), and so on.

I now turn to task of finding fault with the argument against neofregeanism in section three. Here it is again:

> (1) (HP)'s success condition is at least that there exist infinitely many things. (From the reference-based theory of definition)
>
> (2) (HP) can be mathematically basic and knowable *a priori* only if it is knowable *a priori* and independently of (HP) and other mathematical claims that (HP)'s success condition is met. (Premise)
>
> (3) (HP) can be mathematically basic and knowable *a priori* only if it is knowable *a priori* and independently of (HP) and other mathematical claims that there exist infinitely many things. (From (1), (2))
>
> (4) It cannot be known *a priori* and independently of (HP) and other mathematical claims that there exist infinitely many things. (Premise)
>
> (5) (HP) cannot be mathematically basic and known *a priori*. (From (3), (4))

Now I suggested in the last section that 'There exist infinitely many things'[22] might be true in the 'new' post-(HP) language and false in the 'old' pre-(HP) language. This means that the statements that compose this argument are relevantly ambiguous. Let's first consider the argument on the assumption that it is written in the old language. Then, I suggest, premise (1) is false. Even if 'There are infinitely many things' is false in the old language, (HP) can succeed as a definition provided that the modal conservativeness condition is met. Now consider the argument on the assumption that it is written in the new language. Then perhaps (1) is true, for in

---

is conservative one must find a model for it in, say, ZFC.

[22]Or the formal counterpart of this statement in whatever language is used for rational reconstruction.

the new language 'There are infinitely many things' is a trivial definitional truth, and thus arguably its truth is required for the success of any definition (in the same way that 'All bachelors are unmarried' is required to be true for any definition whatever to succeed). So premise (1) seems safe. However, this time I think premise (4) can be rejected. The agent can always establish that (HP) is true by arguing, in the old language if need be, that the modal conservativeness condition obtains; in doing so he need not appeal to any other body of mathematical theory and he need not assume (HP) itself.

I've said that I'm going to be noncommittal about the epistemology of logic in this paper, but I don't think I can avoid the issue entirely. To finish the section, I'll consider the following objection to my position:

> You claim that in order to know that (HP) is true one must know independently that the modal conservativeness condition is met, and this involves knowing that (HP) is consistent. Now to prove that (HP) is consistent requires a substantial body of mathematical knowledge. If 'consistent' means 'has a model', then proving consistency will involve proving that (HP) has a model—and such a proof would require substantial mathematical assumptions. Showing that (HP) is proof-theoretically consistent would also require a lot of mathematics. So either way, one can't know (HP) independently of other mathematical claims after all.

This assumes that the only way of knowing that (HP) is consistent is by *proving* its consistency. I deny this. I think that one can establish that (HP) is consistent without having such a proof. To repeat, I think that one can learn that the modal conservativeness requirement is met by trying unsuccessfully to derive contradictions from one's definitions together with co-possible empirical claims, by investigating (HP)'s consequences, trying to get a clear picture of what a 'model' (in an informal

sense) for (HP) would be like, by investigating statements logically equivalent to (HP), and so on.

3.8 An alternative solution to the infinity problem?

Having presented my response to the infinity problem, I would like to criticise an alternative approach: the 'externalist' approach, as I will call it. The idea is to attack the argument given in section three at its second premise:

> (2) (HP) can be mathematically basic and knowable *a priori* only it is knowable *a priori* and independently of (HP) and other mathematical claims that (HP)'s success condition is met.

The proponent of the externalist approach thinks that this is not right. The idea is that when one uses (HP) as a definition,[23] one thereby knows that (HP) is true, at least in the absence of a defeater. (HP) can be treated as 'innocent until proven guilty', as it were. Having learned (HP), one can confirm retrospectively that its success condition was met.

I argued that this is unattractive back in section three, by pointing out that the analogous position with regard to 'Goliath' is ridiculous. Recall the definition:

> Goliath is a man at least 10cm taller than every other man.

The success condition for this definition, uncontroversially, is that some man is at least 10cm taller than every other man. It would be crazy to say that, provided that such a person exists, a user of this definition is justified *a priori* in believing that this is so.

---

[23] . . . or (HP^3), or whatever—I omit this qualification from now on.

This example refutes the following claim:

> (Strong Externalism) If an agent uses a definition δ, then the agent knows that δ is true, so long as she doesn't have any reason for thinking that its success condition is not met.

It might be replied, however, that the definition of 'Goliath' is relevantly different to (HP). Perhaps the externalist approach is right in the case of (HP) and wrong in the case of 'Goliath'. To make this plausible, it seems to me, the proponent of the externalist approach will need to say something about what the difference is between the two cases. She would need to specify a weak form of externalism, presumably something of the following form:

> (Weak Externalism) If an agent uses a definition δ that meets condition C, then the agent knows that δ is true, so long as she doesn't have any reason for thinking that its success condition is not met.

The challenge is to specify a suitable constraint. One idea would be to say that the definition's success condition must be a necessary truth; alternatively, one could require that the definition satisfy some conservativeness requirement. Either one of these would indeed imply that externalism applies to (HP) but not to the definition of 'Goliath'.

Philip Ebert and Stewart Shapiro have shown that this approach won't work.[24] To see the problem, imagine that an agent seeks to define the vocabulary of number theory by stipulating the conjunction of (HP) with (say) Fermat's Last Theorem. It seems to me that for any reasonable choice of constraint C, it will turn out that this definition meets the constraint, and so Weak Externalism will imply that the agent knows that his definition has succeeded. He will then know that Fermat's Last

[24]See Ebert and Shapiro (2009).

Theorem is true. Now this might be an acceptable conclusion in some cases; perhaps, for example, the agent has already proved that (HP) entails the theorem. But surely in general one cannot learn the truth of mathematical theorems in this way. If only mathematics were so easy!

My own proposed solution has no such unwanted consequence. On my view, one knows that one's definition has succeeded only if one has some (independent) reason for thinking that a suitable interpretation function exists for the newly extended language. It is hard to see how one could have reasons to believe this in the case just imagined without proving that Fermat's Last Theorem is consistent with (HP)—no easy task.

3.9 Some points of disagreement

3.9.1 The importance of abstraction principles

(HP) is an abstraction principle;[25] that is, it has the form:

$$\forall F \forall G(\Omega x : Fx = \Omega x : Gx \leftrightarrow \Gamma(F, G))$$

where Ω is a singular-term-forming operator, and Γ is an equivalence relation on concepts.

Hale and Wright put a lot of emphasis on the fact that (HP) is an abstraction principle; they think that abstraction principles are particularly well-suited to serve as implicit definitions. Hale has developed a version of real arithmetic using abstraction principles as axioms, and they think the project of developing a version of set theory based on abstraction principles is important. They even go as far as to call their position 'abstractionism'. On my account, abstraction principles have no such special

[25] More specifically, a 'conceptual' abstraction principle—see Fine (2002) for an explanation of this term.

status: any definition will succeed, so long as the modal conservativeness condition is met, regardless of its logical form. On my view, one could implicitly define set-theoretic terms simply by stipulating the axioms of (say) ZFC together with some appropriate metaphysical stipulations about sets.

Hale and Wright defend their idea that abstraction principles have a special status in Hale and Wright (2009a). Specifically, they discuss whether the axioms of Dedekind-Peano arithmetic could play the role that they claim for (HP)—whether these axioms could define the relevant terms and provide the basis of knowledge of the truth of the theorems of arithmetic. They begin the discussion by imagining someone who attempts to stipulate the truth of the Ramsey sentence corresponding to the Dedekind-Peano axioms; in effect, they say, such a person says, 'Let there be an omega-sequence!'. Doing this, they point out, is not sufficient to give someone a conception of what numbers are, or to enable singular thought about particular numbers. They then think about what the difference would be between stipulating the Ramsey sentence for the Dedekind-Peano axioms, and just stipulating the axioms themselves. They say:

> If it is fair to characterise the stipulation of the Ramsey sentence as, so to say, the issuing of an injunction
>
> > "Let there be an omega-sequence!"
>
> then it looks as though all that gets added when what is stipulated is not the Ramsification but the second-order Dedekind-Peano axioms themselves is the extra content conveyed by the injunction:
>
> > "Let there be an omega-sequence whose first term is zero, whose every term has a unique successor, and all of whose terms are natural numbers!"
>
> And the trouble is, evidently, that it is not clear whether there really is any extra content—whether anything genuinely additional is conveyed by the uses within the second injunction of the terms "zero", "successor" and "natural number". After all, in grasping the notion of an omega-sequence in the first place, a recipient will have grasped that there will be a unique first member, and a relation of succession. He learns nothing substantial by being told that,

> in the series whose existence has been stipulated, the first member is called "zero" and the relation of succession is called "successor"—since he does not, to all intents and purposes, know which are the objects for whose existence the stipulation is responsible. For the same reason, he learns nothing by being told that these objects are collectively the "natural numbers", since he does not know what natural numbers are. Or if he does, it's no thanks to our stipulation.

They conclude:

> ...the stipulation of Dedekind-Peano, even if the vehicle is assumed to be a necessary truth, conveys no conception of the sort of thing that zero and its suite are—they could be anything at all, provided they are countably infinite and (therefore) allow of a serial order.

They add that the Dedekind-Peano axioms fail to 'communicate a singular-thought-enabling conception of the sort of objects the natural numbers are.' On the other hand:

> Someone who takes it that [(HP)] is true should take himself to have learned that the referents of the newly introduced terms are invariances under one-one correspondence and hence, whatever else may be true of them, effectively provide a measure of that property of a concept which is fixed by its relationships of one-one correspondence to other concepts—its cardinality. [(HP)] thus contributes a characterisation of the nature of (finite) cardinal number that is unmatched by Dedekind-Peano, which convey no more than the collective structure of the finite cardinals—something which, since it entails those axioms, [(HP)] also implicitly conveys. If moreover the stipulation is received as a characterisation of a criterion of identity for the objects concerned, then the effect (or so we have repeatedly argued, modulo Caesar issues) is to convey a sortal concept of number and thereby to provide the means for basic individuative thought of particular numbers.

I agree that the stipulation of the Dedekind-Peano axioms alone does not give one an adequate conception of number, for these axioms do not allow one to determine whether zebras are numbers, or whether Julius Caesar is a number. However, as I have said, (HP) is the same in this respect. If however, one supplements the Dedekind-Peano axioms with some further stipulations (numbers exist necessarily, they have

no location in space and time, ...) one will be in a position to determine the truth-values of these statements, together with purely arithmetical identity statements like '7+5=12'. One can also derive from these stipulations true statements about what properties numbers have. One can learn, for example, that two is prime, and not a Roman emperor. What more is there to having singular thoughts about numbers, or knowing what numbers are?

My guess is that Hale and Wright would respond by rejecting my approach to the Julius Caesar problem, saying perhaps that it is 'theft' simply to make stipulations about the metaphysical characteristics of numbers, whereas to do derive statements about these characteristics from the sortal exclusion principle is 'honest toil'. I cannot see a relevant difference here. Perhaps the difference is supposed to be that (HP) entails that numbers are 'invariances', while the Dedekind-Peano axioms do not have this implication. But how are we supposed to know what 'invariances' are?—this is just pushing the problem around.

The truth based theory of definition that I have developed here implies that one is free to make these stipulations (provide that the modal conservativeness constraint obtains). If Hale and Wright think otherwise, they need to present an alternative to my version of the truth based theory.

3.9.2 Quantifier variance

In their Hale and Wright (2009b), Hale and Wright criticise a philosophical position called 'quantifier-variance':

> Quantifier-Variance is the doctrine that there are alternative, equally legitimate meanings one can attach to the quantifiers—so that in one perfectly good meaning of 'there exists', I may say something true when I assert 'there exists something which is composed of this pencil and your left ear', and in another, you may say something true when you assert 'there is nothing which is composed of that pencil and my left ear'.

I think that they should accept quantifier-variance.[26] I have already explained why Hale and Wright should be open to the idea that the stipulation of (HP) changes the meanings of the quantifiers (see the beginning of section six). I also think that Hale and Wright should say that the 'old' and 'new' quantifiers are 'equally good'—or, better, that the old quantifiers are neither better nor worse than the new ones (what do such comparisons mean, anyway?)[27]—If Hale and Wright claim that the old quantifiers are better, this looks like a rejection of their Platonism; if they claim that the new quantifiers are better, they will then be stuck with the problem of explaining how they could know this (rather bizarre) truth.

So I will now respond to Hale and Wright's criticisms of quantifier-variance.[28] They begin:

> The quantifier-variantist owes us two things: he needs to explain why the allegedly different quantifiers which can all be expressed by the words 'there are' are all quantifiers; and he needs also to tell us how they differ in meaning. The first requires him to identify a common core of meaning for the quantifier-variants; the second requires him to tell us, in general terms, what the variable component is—what the dimension of meaning-variation is.

Both tasks, they think, are impossible. Regarding the first, they suggest the 'obvious answer', that all the quantifiers 'share the same inferential behaviour—are subject to the same inference rules'. This, they say, is 'unsustainable'. To explain why, they suppose that there are two existential quantifiers $\exists^1$ and $\exists^2$ such that $\exists^1 x A(x)$ and $\exists^2 x A(x)$ differ in truth value; then:

> Suppose $\exists^1 x A(x)$. Assume $A(t)$ for some choice of 't' satisfying the usual restrictions. Then by the introduction rule for $\exists^2$, we have $\exists^2 x A(x)$ on our second

[26] Sider (2007) makes the same point.

[27] Actually, I think it would be better for Hale and Wright to say that the new quantifiers are better from a pragmatic point of view, and then question the idea that there is any other appropriate mode of assessment. See paper five.

[28] The points I make below are not original to me. See for example Sider (2007).

> assumption and so, by the elimination rule for $\exists^1$, can infer $\exists^2 x A(x)$ discharging that assumption in favour of the first. We can similarly derive $\exists^1 x A(x)$ from $\exists^2 x A(x)$. Yet by hypothesis, one of the two is true, the other false. It follows that either the inference rules for $\exists^1$, or those for $\exists^1$, are *unsound*—and hence that that one set of rules or the other must fail to reflect the meaning of the quantifier it governs.

They then suggest that there is no way of explaining what the different quantifiers have in common.

I think the 'obvious answer' is in fact quite correct. The problem that Hale and Wright describe doesn't arise if the quantifiers are safely separated in different languages. One can even have two different existential quantifiers in one language if the language is multi-sorted and appropriate restrictions are put on the introduction and elimination rules for the quantifiers.

Hale and Wright also argue that the proponent of quantifier-variance will be hard-pressed to explain the variation in meaning between the different quantifiers:

> As regards the second, it remains very difficult to see how the relevant dimension of variation could be other than the range of the bound variables (or their natural language counterparts)—so that (relevantly) different quantifier meanings differ just by being associated with different domains. But while this answer seems unavoidable, it seems in equal measure unfit for the intended purpose. For ... the quantifier variantist's allegedly different quantifiers can't differ by being different restrictions of some other, perhaps unrestricted, quantifier—for then they wouldn't all be 'equally good'.

I think Hale and Wright are correct that this isn't a good way to think about the variation in meaning between different quantifiers. I think it is better to say, flat-footedly, that they mean something different because sentences containing them have different truth conditions.

Hale and Wright have one further objection. They don't see how you could 'expand' your quantifier if quantifier-variance were true:

> The only obvious suggestion—that by introducing concepts of new kinds of objects (e.g. mereological sum, or number) we somehow enlarge the domain—is,

> in so far as it's clear, clearly hopeless. We cannot expand the range of our existing quantifiers by saying (or thinking) to ourselves: 'Henceforth, anything (any object) is to belong to the domain of our first-order quantifiers if it is an F (e.g. a mereological sum)'. For if Fs do not already lie within the range of the initial quantifier 'anything', no expansion can result, since the stipulation does not apply to them; while if they do, then again, no expansion can result, since they are already in the domain.

Hale and Wright are correct of course that one cannot 'enlarge the domain' in this way, but this doesn't mean that one cannot enlarge the domain at all. One can do so, for example by saying 'Let (HP) be true!'—thereby ensuring that numbers are included in one's domain of quantification.

3.9.3 Conservativeness vs. Irenicity

I have suggested that any definition which satisfies my modal conservativeness requirement will succeed. There is a standard objection to positions like this: it is possible to find pairs of definitions δ_1 and δ_2 such that in both cases the modal conservativeness condition will be met, but δ_1 and δ_2 are inconsistent, because they impose inconsistent constraints on how many things there are. It is even possible to find cases like this where both δ_1 and δ_2 are abstraction principles. To avoid unnecessary technicality, I won't give examples: instead, I refer the reader to Weir (2003) for details. The objection then goes like this:

> Suppose we adopt both δ_1 and δ_2 as definitions. Then on your view, both definitions will succeed, because in both cases the modal conservativeness condition obtains. So both δ_1 and δ_2 will become true. But this is absurd, since δ_1 and δ_2 are inconsistent!

It's worth separating some cases here. Suppose first of all that someone introduces both of these two definitions at once. It is clear, on the truth based account, that the definitions will fail. If there were a suitable interpretation function J^+, both of

the two abstraction principles would have to be true relative to J^+ (by condition (C2*)). But then all the entailments of the two principles would have to be true too (by condition (C3)). Since the principles are inconsistent, they entail everything, and so it would follow that, for example, 'Caesar is a horse' would be true relative to J^+, contrary to (C1*). So in fact no function exists meeting the three requirements.

Now consider a case in which someone adopts δ_1 first, and then subsequently adopts δ_2. Because (we are supposing) the first abstraction principle is modally conservative together with the persons' other definitions, the stipulation will be successful and δ_1 will be true in the person's newly extended language. Then the agent adopts δ_2. δ_2 is inconsistent with a definition already in the agent's language—namely δ_1. It is thus not modally conservative, and the stipulation fails. No problem is generated for the account. The same thing happens if the person introduces the two definitions in the opposite order.

What these cases show is that, according to the truth based account, the legitimacy of a definition may depend on what definitions the person has already made, and that definitions that are legitimate one-by-one may not be jointly legitimate.

In the quotation that follows, Wright criticises this approach. Recall, that δ_1 and δ_2 are inconsistent because they impose inconsistent constraints on how many things there are. Wright realises that on my approach the question 'How many things are there?' will have different answers in different languages. Someone who uses δ_1 as a definition will give a different answer to this question to someone who uses δ_2. Wright thinks that this is not in the spirit of neofregeanism:[29]

> The idea ... was to be that Hume's Principle should be viewable merely as fixing truth conditions of statements about numbers, who satisfaction is then left for determination by how relevant matters independently and objectively stand. When Hume's Principle is laid down, that is to say, as analytic of the

[29]Wright (2001b).

> concept of number, nothing is supposed to be happening inconsistent with the general picture that thinking can invent concepts under which objects may or may not fall but that it cannot in general invent the objects which do or not fall under them. The sortal concepts we actually choose to employ may be only one among many possible groups; but if others are possible, then—the picture is—there already are, or are not, objects falling under them, whether or not we ever came to employ those concepts. The cost of [the view in question] is that it becomes impossible to sustain this kind of thinking about the abstract realm. Rather, we shall have to say that how many objects there are, and hence which objects of which kinds there are, is something that is relative to the scheme of concepts we happen to employ; so that in the abstract realm, our adoption of a particular conceptual scheme affects not merely which objects we shall recognise to exist, as in the concrete case, but which objects actually exist. This is not perhaps an incoherent view; at least, it will not be the work of an instant to show it incoherent. But it is utterly foreign to the Fregean spirit which the new logicism was supposed to safeguard.

I am quite happy to defer to Wright on the question of whether my proposal is 'Fregean'. He may well be right that it is not. However, there is an inference that Wright makes here which baffles me:

> According to the proposed view, the question 'How many abstracta are there?' will have different answers in different languages.
>
> *Therefore:*
>
> According to the proposed view, we 'invent' (i.e. create) abstracta.

I can see no connection between the premise of this argument and its conclusion.

Wright's own criterion for distinguishing the acceptable abstraction principles from the 'bad company' is the 'Irenicity Criterion'. We define 'Wright-Conservative' as follows:

> Given a functor Ω defined by an abstraction principle A, let P be a predicate defined by $\forall x(Px \leftrightarrow \neg\exists F(x = (\Omega y : Fy)))$. Then the abstraction principle A is Wright-Conservative over a base theory T iff whenever $T^P \cup \{A\}$ implies some statement α^P, T alone implies α.

'Irenic' can then be defined by saying that an abstraction principle A is Irenic just in case:

(i) A is Wright-conservative over any base theory.

(ii) A is consistent with any other conceptual abstraction principle which meets criterion number (i).

I hope to have convinced you that my own 'Modal Conservativeness' condition has a solid theoretical rationale. If so, then it is preferable to Wright's Irenicity requirement, which is motivated only by inspection of cases.

3.10 Appendix: The modal conservativeness criterion

In this appendix, I will demonstrate that an interpretation function that meets constraints (C1*), (C2*) and (C3) will exist provided that the modal conservativeness requirement is met.

Before proceeding to a demonstration, I should pause to justify an important premise that I will need:

Every set of statements Γ has a maximal consistent superset Γ^+.

I'll defend this assumption in two cases: first, there's the case in which the agent has a semantic notion of entailment; second, there's the case in which the agent has a syntactic notion of entailment. See section five for some clarification of this. In the first case, it's trivial that every consistent set has a maximal consistent superset. Consistency, in this case, just amounts to having a model. If Γ is a consistent set, it has a model $\mathcal{M}$. Let Γ^+ be the set of all statements true at $\mathcal{M}$. Then Γ^+ is a maximal consistent superset of Γ. In the second case my assumption is just Lindenbaum's Lemma, which can be established in the familiar way. I'm ignoring systems that allow for infinitely long proofs.

Now I'm ready to show that an interpretation function that meets constraints (C1*), (C2*) and (C3) will exist provided that it meets the modal conservativeness requirement. To begin, let's suppose that a language L with interpretation function J is extended by way of definitions $\delta_1, \ldots, \delta_m$, and suppose that $\delta_{m+1}, \ldots, \delta_n$ are the agent's existing definitions. Let L^+ be the new, extended language. Suppose also that the agent takes it that a predicate P in the newly extended language delimits the 'old' ontology.

Now suppose that the modal conservativeness condition is met. That is, suppose that:

> For any possible world w, $\{\delta_1, \ldots, \delta_n\} \cup \{\alpha^P \in L^+ \colon w \in J(\alpha)\}$ is consistent.

For each possible world w, let T_w be a maximal consistent superset of $\{\delta_1, \ldots, \delta_n\} \cup \{\alpha^P \in L^+ \colon w \in J(\alpha)\}$. Then define J^+ as follows:

> For any $\phi \in L^+$, let $J^+(\phi) = \{w \colon \phi \in T_w\}$

It remains to check that J^+ meets the three conditions. Here they are, for ease of reference:

(C1*) For any 'old' sentence ϕ, $J^+(\phi^P) = J(\phi)$.

(C2*) For any definitional truth δ (either one of the new definitions, or an 'old' definition), $J^+(\delta)$ is the set of all possible worlds.

(C3) J^+ respects entailment, in the sense that if w is an element of $J^+(\gamma)$ for each γ in some set Γ, then w is an element of $J^+(\alpha)$ for any α entailed by Γ.

I'll look at them in turn.

(C1*) Assume that ϕ is an 'old' sentence. We want to show that $J^+(\phi^P) = J(\phi)$.

"$\subseteq$" Suppose $w \in J^+(\phi^P)$, so $\phi^P \in T_w$. Now suppose for *reductio* that $w \notin J(\phi)$. Then assuming that J assigns truth conditions in a consistent fashion, $w \in J(\neg\phi)$. Then $\neg\phi^P \in \{\alpha^P \in L^+ \colon w \in J(\alpha)\}$, so $\neg\phi^P \in T_w$—which is a contradiction since T_w is consistent. So in fact $w \in J(\phi)$, as required.

"$\supseteq$" Suppose $w \in J(\phi)$. Then $\phi^P \in \{\alpha^P \in L^+ \colon w \in J(\alpha)\}$, so $\phi^P \in T_w$. Hence, $w \in J^+(\phi^P)$.

(C2*) For any possible world w, T_w is a superset of $\{\delta_1, \ldots, \delta_n\}$ so for any definitional truth δ, $\delta \in T_w$. It follows that $w \in J^+(\delta)$. Therefore $J^+(\delta)$.

(C3) Suppose that w is an element of $J^+(\gamma)$ for each γ in some set Γ, and let α be some statement entailed by Γ. We want to show that w is an element of $J^+(\alpha)$.

Suppose for *reductio* that $w \notin J^+(\alpha)$. Then $\alpha \notin T_w$. Then since T_w is maximal consistent, $\neg\alpha \in T_w$. Now $w \in J^+(\gamma)$ for each $\gamma \in \Gamma$, so $\Gamma \subseteq T_w$. Then $\Gamma \cup \{\neg\alpha\} \subseteq T_w$, so T_w is inconsistent. This is a contradiction, so in fact $w \in J^+(\alpha)$, as required.

Chapter 4

Analyticity in Mathematics, Part Two

4.1 Introduction

To speak loosely, '$Nx : \ldots x \ldots$' is an operator which means 'the number of things x such that $\ldots x \ldots$'. For example, '$Nx : (cat(x) \wedge black(x))$' refers to the number of black cats.

The operator was invented by Crispin Wright, in Wright (1983), who stipulated that this formula, 'Hume's Principle', is its definition:

$$\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \sim G)$$

Here, '$F \sim G$' abbreviates a formula which means 'there is a one-to-one correspondence between the Fs and the Gs'.[1] So Hume's Principle as a whole means roughly that for any concepts F and G, the number of Fs is the number of Gs just in case

[1]To be more precise about it, '$F \sim G$' abbreviates:

$$\exists R[\forall x(Fx \rightarrow \exists y(Gy \wedge \forall z(Rxz \leftrightarrow z = y))) \wedge \forall y(Gy \rightarrow \exists x(Fx \wedge \forall z(Rzy \leftrightarrow z = x)))].$$

there is a one-to-one correspondence between the Fs and the Gs.

Having defined the '$Nx : \ldots x \ldots$' operator in this way, Wright used it to explicitly define some other number-theoretic terms. For example, he defined '0' like this:

$$0 := Nx : x \neq x$$

He also gave definitions of 'natural number' and 'predecessor' using his new operator (see Wright (1983) for details). He then deduced the second-order Peano axioms from these definitions.

Wright claimed that Hume's Principle, being a definition, is an *a priori* truth. And he said that the definitions of '0', 'natural number' and 'predecessor' are *a priori* too. So by deducing the Peano axioms from these definitions, Wright claimed to have achieved *a priori* knowledge of the truth of these axioms.

In later work, Wright and his collaborators (the 'neofregeans') have developed a detailed defence of the claim that Hume's Principle is a legitimate definition, and hence an *a priori* truth. They have also attempted to extend Wright's approach to other branches of mathematics. Bob Hale has developed a neofregean treatment of real analysis, in Hale (2001). More ambitiously, the neofregeans are working on set theory.

In this paper, my question is, 'What should neofregeans say about the mathematical knowledge of ordinary working mathematicians, engineers and scientists?' These people know a great deal of mathematics—but apparently they didn't get to this knowledge by the neofregean route. The vast majority of mathematicians have not heard of the '$Nx : \ldots x \ldots$' operator or of Hume's Principle, and yet they know that 17 is prime. And the vast majority of engineers have not seen Hale's treatment of real arithmetic, but they know that $\sqrt{2}$ is irrational. And most mathematicians know a good deal about sets, even though a neofregean set theory has yet to be developed.

According to the neofregeans, in Wright's version of Peano arithmetic all the non-logical words have definitions, and all the theorems are analytic, in the sense that

they are entailed by definitions.[2]

Neofregeans might be attracted by this generalisation of this claim:

Definitionalism

(D1) For any person S at time t, all of S's non-defective mathematical words have definitions.[3]

(D2) For any person S at time t, if a sentence is (part of) a definition of a non-defective word for S at t, that sentence is *a priori* for S at t.[4,5]

(D3) For any person S at time t, if any purely mathematical sentence is knowable for S at t, then that sentence is entailed by sentences which are definitions for S at t.

This position was defended by some of the logical positivists: see for example Ayer (1936). It became unpopular after Quine's famous discussion of the analytic/synthetic distinction, but Quine's critique has not the sway that it once had.[6] In this paper

---

[2]This is meant as a stipulation. In this paper, a sentence is analytic just in case it is entailed by definitions. It has been suggested that 'analytic' should be understood more broadly. For example, 'Anything yellow is coloured' doesn't seem to be entailed by definitions, but perhaps it should be counted as analytic nonetheless. As far as I know, it has never been claimed that mathematical theorems or axioms are analytic but not entailed by definitions (indeed, Hale and Wright explicitly reject this suggestion in Hale and Wright (2000)) so I set aside the issue.

[3]If the qualification 'non-defective' were omitted, counterexamples to (D1) would be easy to find—for example in the works of 'mathematical cranks' (see Dudley (1992), particularly the chapter 'Incomprehensibility of Crank's Works'). I won't discuss what 'defective' means, since I take it that the term is clear enough for my purposes.

[4]Arguably, there are some mathematical definitions which are not true. For example, suppose I define the notation
'$\{x : \ldots x \ldots\}$' like this:

$$\forall F \forall G(\{x : Fx\} = \{x : Gx\} \leftrightarrow \forall x(Fx \leftrightarrow Gx))$$

This sentence is presumably not true, even though (arguably) it is a definition. However, this is not a counterexample to (D2) since, having been defined in this faulty way, the '$\{x : \ldots x \ldots\}$' notation is defective.

[5]When I say that a sentence is *a priori*, I mean that it is *a priori* that the sentence is true.

[6]Quine's critique is Quine (1953). Boghossian (1996) is an influential critique of Quine's position.

I will argue that, whatever the deficiencies of Quine's argument, his conclusion was correct.

My objection to definitionalism starts with the question, 'What do you mean by 'definition'?' It seems to me that definitionalism is not an attractive position in the absence of a good answer to this question. I'll look at a number of different answers that definitionalists might offer, and argue that none of them is adequate. More precisely, for each interpretation of 'definition' I will argue that (D1), (D2) and (D3) are not all true.

4.2 Definition by private stipulation

This is perhaps the simplest way of drawing the distinction between definitions and sentences. The idea is that each of us is in charge of her own idiolect. A sentence in a person's idiolect is a definition just when that person has *stipulated* that it is a definition.

This version of definitionalism fails because, on this interpretation of 'definition', (D1) and (D3) are false. With very few exceptions, only professional pure mathematicians bother to make stipulations about which mathematical sentences are definitions. The typical scientist, for example, makes no such stipulations. It follows, given this interpretation of 'definition', that no sentence is a definition for the typical scientist. We should not conclude that the typical scientist has no non-defective mathematical words, and no mathematical knowledge.

4.3 Definition by expert consensus

A definitionalist might see this as an illustration of the familiar point that language is a social phenomenon. Individuals, she might say, don't usually need to make stipulations about what their mathematical words are to mean, because these words

are of a shared, public language. Definitionalists, it might be said, should be *social externalists*. In this section, I'll discuss a social externalist version of definitionalism, inspired by Putnam's discussion of 'the division of linguistic labour' in Putnam (1975).

The proposed account of definition is this: a sentence is a definition just in case there is a consensus among the experts of the relevant field that it is a definition. I have two objections to the version of definitionalism which uses this account of definition.

Here's the first. If you hunt around through maths textbooks, you will find that the definitions they contain are in many cases not consistent. The different writers have different audiences, and different tastes, and so on, and so they choose different definitions.

Examples are easy to find, even if we consider only the real line:

- In some books, 'natural number' is defined in such a way that its extension includes zero; in other books the predicate is defined more narrowly, so that it excludes zero.

- In books on pure mathematics, numerical terms are often introduced by 'set-theoretic construction', and the various constructions that are used are not consistent, as Paul Benacerraf famously pointed out, in Benacerraf (1965).

- In many mathematics textbooks, the definitions of 'natural number', 'integer', 'rational number', and 'real number' are such as to imply that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. For example, 'rational number' might be given the definition:

  > A real number q is rational iff for some integers m and n, qm=n.

  On the other hand, when these different kinds of number are introduced by set-theoretic construction, these inclusions are usually violated. For example,

> if real numbers are identified with Cauchy sequences of rationals, then $\mathbb{Q} \subset \mathbb{R}$ will fail.

A consequence of all of this inconsistency is that (D1) and (D3) are false, on the current understanding of definition:

> (D1) For any person S at time t, all of S's non-defective mathematical words have definitions.
>
> (D3) For any person S at time t, if any purely mathematical sentence is knowable for S at t, then that sentence is entailed by sentences which are definitions for S at t.

As we have seen, there is no expert consensus about how the words of (for example) number theory are defined. On the suggested understanding of 'definition', it follows that the words of number theory lack definitions. Assuming (D1), we would have to conclude, absurdly, that these words are defective. And given (D3), we would have to conclude, again absurdly, that none of us have any knowledge of number theory.

I introduce my second objection to the 'expert consensus' version of definitionalism with an example. Consider the empty set symbol, '$\varnothing$'. This symbol with its current meaning was introduced by the Bourbaki group in a book they published in 1939.[7] Now consider the members of the group working on the draft of the book, prior to publication. At this time, '$\varnothing$' had no definition by expert consensus—since only a small number of experts knew of the new symbol. And yet the new symbol was not defective. Moreover, consider the theorems in the manuscript which contained the symbol '$\varnothing$'. These theorems were not entailed by definitions by expert consensus, and yet surely the authors of the book knew them to be true. The point more generally is that the proponents of this version of definitionalism are unable to account for the

[7]The book is Bourbaki (1939). For details of how the '$\varnothing$' symbol was invented, see Weil (1992).

fact that sometimes one knows a mathematical sentence to be true, even when that sentence contains terminology that is not widely known among experts.

4.4 Kripkean definition

In this section I consider an alternative social externalist account of definition, based on Kripke's discussion of proper names in Kripke (1980). The idea is that it is the person who *coins* a word who gets to choose its definition—the word then retains its stipulated definition as it is passed from person to person. Consider for example the term 'perfect', as it is used in number theory. Presumably someone in the past stipulated that 'perfect' is to be defined by:

> A number is perfect just in case it is equal to the sum of its proper positive factors.

The word 'perfect' has then been passed from person to person, and of course it is now widespread.

Before getting to my objection, I would like to discuss a different objection which I *don't* take to be decisive. Most mathematicians only apply the word 'definition' to *explicit* definitions.[8] It follows (so the objection goes) that in mathematics, the only 'Kripkean' definitions are explicit definitions. But of course definitionalists maintain that some definitions are implicit. I think that the argument is not decisive, because the definitionalist can respond like this:

> For historical reasons, mathematicians reserve the term 'definition' for *explicit* definitions; they call the *implicit* definitions 'axioms'. But despite this confusing terminology, mathematicians do think of the axioms of set

[8]There are exceptions—see Hilbert (1899). But these are so rare that I think they can be safely ignored.

> theory (for example) as fixing the meanings of those set-theoretic words which lack explicit definitions. So plausibly these axioms are Kripkean definitions.

As far as I can see, this is the definitionalist's *only* reasonably plausible response to the objection. If I'm right about this, the definitionalist who makes use of a Kripkean account of definition will have to commit herself to the claim that all the definitions in mathematics are sentences that have been labelled 'definition' or 'axiom' by a mathematician.[9]

Now to an objection to this version of definitionalism which *is* decisive.

Early axiomatisations of real arithmetic did not include a completeness axiom, and in consequence they were too weak. They failed to entail, for example, that 2 has a square root. For example, in his landmark 1821 textbook *Cours d'Analyse*,[10] Cauchy did not include a completeness axiom. In consequence, some of his proofs were flawed. For example, the intermediate value 'theorem' was not entailed by Cauchy's axioms, and his 'proof' of it goes blurry at exactly the point where a completeness axiom is needed—see Lützen (2003). A correct proof of the intermediate value theorem, in one of today's textbooks, will make ineliminable use of a completeness axiom. Given that no completeness axiom had been formulated at the time, no completeness axiom was a Kripkean definition back in 1821. Indeed, no axiomatisation of real analysis at the time was strong enough to entail the completeness of the real line. And so the intermediate value theorem was not entailed by Kripkean definitions back in 1821. But surely Cauchy *knew* the intermediate value theorem to be true. The intermediate value theorem is *obviously* true. This is a counterexample to (D3), given the Kripkean

---

[9]For reasons that I need not discuss here, Hale and Wright have explicitly rejected the idea that one could fix the meanings of '0', 'natural number' and 'successor' by using the Peano axioms as stipulative definitions, see their Hale and Wright (2009a). So they would have to reject the Kripkean version of definitionalism. Still, I think the position worth discussing.

[10]Cauchy et al. (2009).

account of definition:

> (D3) For any person S at time t, if any purely mathematical sentence is knowable for S at t, then that sentence is entailed by sentences which are definitions for S at t.

4.5 Disposition-based accounts of definition

The Cauchy case is a counterexample to the version of definitionalism which employs the 'Kripkean' account of definition. However, it also has a much broader significance: it is a counterexample to *any* version of definitionalism according to which a sentence is a definition only when it has been stipulated that this is so. Quite simply, in 1821 Cauchy knew that the intermediate value theorem is true, even though no set of stipulative definitions available at the time was strong enough to entail it.

So we must consider versions of definitionalism according to which it doesn't take a stipulation to make a definition. In this section, I discuss an alternative approach, inspired by the Cauchy example.

A modern reader of Cauchy's 'proof' of the intermediate value theorem is tempted to say that Cauchy was 'tacitly assuming' that the real line is complete. So here's a suggestion: perhaps by tacitly assuming some completeness axiom, Cauchy gave it definitional status. More generally, perhaps the definitionalist should say that a judgement will be definitional for a person, even in the absence of an explicit stipulation to this effect, *if that person treats it as such.*

In pursuing this line, the definitionalist will have to say something about what it is to 'treat something as a definition'. I suppose there will be several different ways of doing this, but presumably the gist will be that someone treats a sentence as a definition when she does nothing to provide evidence of its truth (beyond perhaps providing reasons for thinking it a legitimate definition), but does assume it in the

course of her reasoning, and in particular is willing to use it as a (stated or tacit) assumption in proofs.

According to this suggestion, whether or not a judgement counts as a definition, for a particular person, depends on that person's inferential dispositions. So I'll call accounts of definition like this 'disposition-based' accounts.

This is the simplest disposition-based account:

> *The Crude Disposition-Based Account*
>
> A sentence α is a definition for a person S just in case S is disposed to treat α as a definition.

It would be a mistake for a definitionalist to adopt this account. Surely, in non-ideal circumstances, someone might be disposed to treat a false sentence as a definition.

Here is a much better theory:

> *The Sophisticated Disposition-Based Account*
>
> A sentence α is a definition for a person S just in case S's dispositions are such that, under ideal conditions, S would treat α as a definition.

A proponent of this account would have to spell out what conditions are 'ideal' in the relevant sense. Presumably when S is under ideal conditions, she would have access to all the resources she needs (books, calculators etc.), her cognitive abilities would be so far advanced that she would make no mistakes when constructing proofs, and so on. But I won't stop to figure out all the details.

I have two objections to this version of definitionalism.

First, this version of definitionalism commits the 'conditional fallacy' (see Shope (1978)). There are plenty of people around who use *clumsily defined* mathematical notations. Presumably, these people would, under ideal conditions, *abandon* their clumsy terms and use better ones instead. Assuming (D3), and the sophisticated

disposition-based interpretation of 'definition', these people don't know the truth of any mathematical sentences that contain the clumsy notation. But we should reject this conclusion.

So as not to introduce any irrelevant arguments about which actual mathematical notations are clumsy, and how they would best be corrected, I'm going to use a hypothetical and rather contrived example.

We can imagine a community of mathematicians who don't have our term 'cardinality' or a synonym. Instead, they use their own term 'schmardinality', which is defined in such a way that finite sets have the same schmardinality just in case they can be put in one-to-one correspondence, while all infinite sets have the same schmardinality.[11] The term 'schmardinality' will be much less useful than our term 'cardinality', so if the mathematicians in our imaginary community were put into ideal conditions they would abandon the former term in favour of a synonym of the latter. In this situation, according to the sophisticated disposition-based theory of definition, the word 'schmardinality' has no definition. Assuming definitionalism, it follows that our imaginary mathematicians doesn't know, for example, that this is true:

> $\mathbb{N}$ has the same schmardinality as $\mathbb{R}$.

But it seems to me that the imaginary mathematician might well know that this sentence is true.

Now for the second problem. There are many terms in mathematics which have

---

[11] 'Schmardinality' might be defined like this:

> A schmardinality is a downward closed subset of $\mathbb{N}$.
>
> A set S has schmardinality at least c just in case there is an injective function $f : c \to S$.
>
> A set S has schmardinality exactly c just in case for any cardinality c', S has cardinality at least c' just in case $c' \subseteq c$.

several equivalent definitions. For example, here are two definitions of 'prime' for the natural numbers:

1. A natural number p is prime just in case there do not exist natural numbers $m, n < p$ such that $p = mn$.

2. A natural number p is prime just in case $p \neq 1$ and for any $m, n \in \mathbb{N}$, if $p \mid mn$, either $p \mid m$ or $p \mid n$.

Now consider an ordinary user of the word 'prime', someone who is not a pure mathematician and who hasn't chosen one of these definitions (or some other definition) as her own. Let's call her 'Suzie'. It may well be that neither of these counterfactuals is true:

> Were Suzie under ideal conditions, she would treat 1 as a definition.
>
> Were Suzie under ideal conditions, she would treat 2 as a definition.

I hope it is intuitively clear that it could be that neither of these counterfactuals is true. This intuition is supported by standard semantic theories for counterfactual conditionals. It could be that in some of the closest possible worlds in which Suzie is in ideal conditions she chooses to treat 1 as a definition, and in others she chooses 2. In this case, semantic theories for counterfactuals of the Lewis/Stalnaker variety imply that neither of the two conditionals above is true (see Lewis (1973) and Stalnaker (1968)).

In this case, then, the sophisticated disposition-based theory of definition will have the consequence that, in the actual world, Suzie's term 'prime' has no definition. But presumably Suzie may still know, say, that 11 is prime. This is a counterexample to (D3), given the current account of 'definition':

(D3) For any person S at time t, if any purely mathematical sentence is knowable for S at t, then that sentence is entailed by sentences which are definitions for S at t.

So neither the crude nor the sophisticated disposition-based accounts of definition are adequate for the definitionalist.

4.6 Conclusion

Whatever the merits of Wright's claim to have developed a version of number theory in which all the theorems are analytic, it is not true *in general* that our knowledge in pure mathematics is knowledge of analyticities. The neofregean epistemology of mathematics is incomplete.

Chapter 5

Against Elitism

5.1 Introduction

Elitists distinguish the *elite*, upper-class words from the lower-class *plebeian* ones. People tell me for example that terms from fundamental physics might be elite, like perhaps 'electron' and 'mass'. Some logical constants might be elite too, like maybe the existential quantifier or the identity sign. Perhaps some mathematical words are elite as well. By contrast, secondary quality words like 'red' and 'sour' are paradigmatically plebeian terms. The same goes for aesthetic terms like 'delicious' and 'elegant'. Folksy, unscientific terms like 'dove' are said to be plebeian too.[1]

The thought here is that the elite words are somehow particularly well-fitted to the world's intrinsic structure, so that using elite vocabulary one can describe, as Bernard Williams put it, the world 'as it really is independently of our thought'.[2] When using the plebeian terms, however, we can at best describe the world as it appears to us, with our own peculiar sense organs, tastes and history.

---

[1]We subdivide the *columbidae* into doves and pigeons, but the division is not in good zoological standing. Roughly, the birds of prettier species are called 'doves' while the ugly ones are called 'pigeons'.

[2]Williams (1978), pg. 196.

Here's a rough and ready classification of the available views. The *elitists* are those who accept the idea that one of the proper goals of pure inquiry is to identify and use elite terms;[3] *egalitarians* reject this claim—most egalitarians reject the elite/plebeian distinction altogether. We can subdivide the elitists into two groups: the *naïve* and the *sceptical*. The naïve are those who think that we are capable of figuring out which the elite terms are; the sceptical elitists deny this. Some of the naïve elitists are enthusiasts for the *elitist project*—the attempt to figure out which the elite terms are. The sceptical elitists, of course, think that this project will not succeed.

I am an egalitarian, and I would *like* to defend egalitarianism in this paper. Sad to say, my argument doesn't get me that far. So I will have to settle for a weaker thesis. I will argue that *if* there is a distinction to be drawn between the elite and the plebeian expressions, we cannot work out which the elite expressions are. So my conclusion will be the disjunction of egalitarianism and sceptical elitism.

I will draw heavily on the work of the elitist Sider,[4] who has done a great deal in recent years to clarify the issues and explain what's at stake. In this paper, Sider is both hero and villain: hero, because he has done such a great job at straightening out the topic; villain, because he's on the wrong team.

5.2 Some elitists

I have a few preliminary things to discuss before I get to the argument. In this section, I'll discuss two of my favourite elitists—Bernard Williams and Sider—if only to show that I'm not attacking a straw man. In section three I'll explain why elitism is important. After that, in section four, I'll give you a more careful characterisation of the doctrine. Then I'll start on the argument in section five.

---

[3]See section four for a more careful characterisation of elitism.

[4]See in particular his Sider (2011).

5.2.1 Bernard Williams

Bees' eyes can respond to ultra-violet light—light that is outside the 'visible' range for humans. This means that bees can see and respond to patterns on flowers which are invisible to people unaided by special cameras. Most birds have four different sorts of cone on their retinas, which means that their colour-space is four-dimensional: unlike our own which has only three dimensions (which invites the question, 'What is it like to be a bird?'). These facts remind us that our colour-classifications are idiosyncratic. Only creatures with visual systems rather like ours would classify things as red, or blue. The same doesn't seem to be true of many scientific classifications. When the intelligent aliens arrive, we shouldn't be surprised if they have a word which means *gold*, but it would be astonishing if they have a word which means *yellow*.

Starting with these uncontroversial observations, Bernard Williams argued that one of the goals of science should be to identify classifications which are not 'peculiar' to us. We should coin predicates corresponding to these classifications, Williams thought, and use this special vocabulary in scientific theorising. Only statements couched in these terms could describe the world as it is 'anyway'.[5]

Williams was not an error-theorist about colour-talk: he would have agreed, for example, that it is true that not yet ripe bananas are yellowy-green. But he would add that such claims somehow fail to be fully revelatory of the true natures of things; to describe something as yellowy-green is to describe it merely as it seems, rather than as it 'really' is. Science should aim to eliminate descriptions like this. The goal should be the 'absolute conception of the reality': a description of the world as it 'really' is.

---

[5]Williams (1978), pg. 48.

5.2.2 Sider

Partly in response to David Armstrong's work on universals,[6] David Lewis suggested that properties can be ordered according to 'naturalness': some properties are more natural than others.[7] The more natural properties are those that 'carve nature at the joints'; objects that share a natural property resemble each other in one respect—not just to us, but objectively. The property of being an electron is very natural; the property of being a lion is less so; the property of being a dove is still less natural; all of these properties are more natural than the property of being either a dove or an aqueduct, or a prime number that isn't seventeen.

Plausibly, one of the goals of science is to identify the natural properties. In zoology, for example, it was an advance when we started categorising whales, dolphins and porpoises with the other mammals rather than with the fish: we discovered a way of categorising the animals that better respects the objective similarities and differences between them. Let's say that a predicate corresponding to a natural property is a 'natural predicate'; then it's plausible to say that good scientific theories are expressed using natural predicates.

Sider generalises this to other parts of speech.[8] Just as in science and metaphysics we aim to use natural predicates, so we should aim to use natural quantifiers, natural singular terms, natural operators, and so on. These expressions 'carve at nature's joints', they reveal 'the structure of the world'.

The idea that there are 'natural' quantifiers will be particularly important in what follows. To borrow Sider's example, we could define the word 'schmexists' by stipulating that 'There schmexists an F' means that the property of being an F is a

---

[6] Armstrong (1978).

[7] Lewis (1983); see also Lewis (1984).

[8] Sider (2011).

property expressed by some predicate in Sider's latest book. Sider does not deny that 'schmexists' is a meaningful expression, or that it is a quantifier, and he would agree that some sentences containing the term are true. For example, it is true that there schmexists a person, since the predicate 'person' appears in Sider's book. But Sider would add that 'schmexists' is a highly 'unnatural' word; it's as if we'd introduced a predicate with the stipulation that it be true of doves, aqueducts, prime numbers other than 17, and nothing else.

5.3 What's at stake?

Now I want to discuss why elitism is important. I want to convince you that my topic matters before I get into the argument.

5.3.1 Elitism and the goals of ontology

Consider the claim that there exists a hole in the middle of every ring doughnut. A metaphysician, considering this claim, might find herself conflicted. She might have some theoretical reason for denying that there are holes. At the same time, philosophical modesty, or respect for common sense, might make her feel uncomfortable rejecting the apparent truism that there exists a hole in the middle of every ring doughnut. Surely, she might think, even children know that ring doughnuts have holes in them.

Elitists have a solution. An elitist could respond by saying that 'There exists a hole in the middle of every ring doughnut' is true in English, but only because English quantifiers are plebeian. The elitist could continue by saying that we can introduce by stipulation a new, elite quantifier 'there really exists'. Having done so, we are in a position to say that 'There exists a hole in the middle of every ring doughnut' is true, but 'There really exists a hole in the middle of every ring doughnut' is false. In this

way, the elitist can preserve common sense while hanging on to a sort of metaphysical eliminativism about holes.

I should probably give you a 'real world' example of this kind of thing. Derek Parfit says that numbers and reasons exist 'non-ontologically'.[9] When I first heard him say this, I thought it was obviously daft. Ontology, I thought, just is the study of what there is, or of what exists. So to say that there are these things, numbers and reasons, and that they exist, but do so *non-ontologically*, struck me as flatly inconsistent.

I don't any longer think that Parfit's position is inconsistent. I now think that Parfit is an elitist. When he puts the word 'ontologically' in a sentence, he means to indicate that he's using his quantifier in the elite way. 'Non-ontologically', on the other hand, signals that he's using a plebeian quantifier.

I should mention that Parfit is not alone in saying things like this. GE Moore said similar things in *Principia Ethica*.[10] Jody Azzouni says that there are numbers and fictional characters, but then claims not to be 'ontological committed' to them.[11] Kit Fine thinks that cities exist, but that this is not a 'real' truth[12]—and so on, and so on. So elitism is pretty common, I think. Sider may be the first to go on record with the claim that an elite quantifier is to be distinguished from lesser, plebeian quantifiers—but the idea predates him (as he is well aware).

If these elitist views are correct, this has an important implication about the goals of ontology. *Pace* Quine,[13] inquiry in ontology should not be centred on the question 'What is there?' or 'What exists?'; rather, the goal should be to answer the question

---

[9]Parfit (2009).

[10]Moore (1903).

[11]Azzouni (2006).

[12]The term 'real' is introduced in Fine (2001). The 'city' example is one that he has used in seminars.

[13]Quine (1948).

'What *really* exists?' or 'What is there *really*?'.

5.3.2 Elitism and realism

Think about the dispute between Platonists and nominalists in the philosophy of mathematics. The dispute centres on the apparent existential commitments of standard mathematics, claims like:

- There exists a number which is not the successor of any number.
- Every set has a power set.
- There exists a prime number greater than a million.

Platonists are 'realists' about claims like this, nominalists are not.

One might attempt to defend Platonism by arguing that these existential claims are 'analytic', in the sense that they can be logically deduced from definitions. There are many objections to this sort of position; of which I want to mention just one, which goes like this:

> Okay, I'm willing to concede that 'there are prime numbers greater than a million' is true by definition—but I don't think that this is a sufficient defence of Platonism. The true Platonist will claim not only there are numbers, but also that there are numbers in the elite sense of 'there are'. I don't doubt that you can gerrymander a sense of 'there are' so as to get a true reading of 'there are numbers': but you need to show that 'there are numbers' is true in the elite sense of 'there are', not in some silly gerrymandered sense.

This illustrates a more general point. When asked to characterise 'realism' about some class of statements, elitists will tend to be more demanding in their definitions.

It might be thought that this is all terminological. Who cares what 'realism' means? It is, after all, a technical term and it's probably ambiguous in any case. I agree that we shouldn't worry too much about who deserves the title 'realist'. What is not merely terminological is the question of how philosophers should spend their time. To take a slightly different sort of example, if elitism is true, then whether or not 'good' is elite is an important question in metaethics. If not, it isn't.[14]

5.3.3 Which metaphysical disputes are substantive?

I've heard people argue about how many oceans there are on Earth. The monists think that there is only one 'world ocean', which encircles the globe. Some people think there are four: the Pacific, the Atlantic, the Indian and the Arctic. Others would include the Southern Ocean, and add that the North Atlantic and the South Atlantic are distinct, reaching a total of 6.

My own view? I'm a 'nonsubstantivist' about the issue. I think that the people on different sides of this dispute have just hit upon different ways of using the word 'ocean'. When the monist says, 'there is only one ocean on earth' and one of his 'opponents' says 'there are four oceans on earth', and another says 'there are six oceans on earth', all parties speak truly. Despite the appearance of disagreement, all three are correct. We can sensibly ask which of these several ways of using the word 'ocean' is more closely in line with ordinary English; we can also sensibly ask which of these ways of speaking is more elegant, or more practical. But otherwise there is nothing to fight about here.

Hilary Putnam has a similar nonsubstantivist attitude about many of the disputes in metaphysics.[15] For example, he is a nonsubstantivist about the debates concerning

[14]To understand the relevance of elitism to metaethics, it's interesting to contrast Fine (2001) with Dworkin (1996).

[15]See Putnam (2004) and Putnam (1988).

mereology. He imagines a dispute between two rival metaphysicans, called 'Carnap' and 'Leśniewski'. The latter adheres to the axiom of unrestricted fusion; the former does not, thinking perhaps that there are no disconnected physical objects. Putnam then imagines a small possible world. As Carnap describes it, the world contains just three physical individuals, which we'll call 'a', 'b', and 'c'. Leśniewski agrees that a, b and c exist at the world, but insists that there are also four other objects: a+b, b+c, c+a and a+b+c.

Now Carnap and Leśniewski seem to make incompatible claims about the little world. Indeed, their claims would seem to be logically inconsistent. For example, Carnap thinks that this is true at the little world:

$$\forall(x = a \vee x = b \vee x = c)$$

While Leśniewski claims on the contrary that *this* is true at the little world:

$$\exists x(x \neq a \wedge x \neq b \wedge x \neq c)$$

Nevertheless, Putnam claims that Carnap's and Leśniewski's descriptions are not in fact incompatible: they have two different ways of describing the same possible world. This is analogous to my claim that the disputants in the argument about oceans are not really offering incompatible accounts of world geography—they are just using words differently.

A simple way of defending this idea is to say that Carnap and Leśniewski speak different languages, which are (at least mostly) intertranslatable. For example, when Carnap says 'There exist three objects', this can be translated into Leśniewski's language as 'There are three connected objects'; and where Leśniewski says 'a+b is entirely red', this can be translated into Carnap's language as 'a and b are both entirely red'.

There are several different ways in which one might object to Putnam's nonsubstantivism about mereology. One response is to deny that Carnap and Leśniewski

could be speaking different languages, in the sort of case that Putnam is imagining. If this is right, then *pace* Putnam at least one of the two theories is false.

Here is a different way of attacking Putnam's position—this is a criticism you might expect from some who accepts the axiom of unrestricted fusion:

> I'm willing to concede that Carnap's claims are true. However, if this is right it could only be because Carnap's quantifiers are restricted. When Carnap says 'everything', what he really means is something like 'Every *connected* thing'. If so, then while Carnap's theory is true, it is inadequate in another respect—in his descriptions of the possible world, Carnap 'misses out' some objects, so his description is incomplete. Carnap is not using an elite quantifier. He can tell us 'what there is' in his reduced sense of 'there is', but he fails to tell us what *really* exists, because he lacks the elite quantifier. Leśniewski's theory is not incomplete in this way: he uses an elite quantifier. In consequence, Leśniewski's theory is superior to Carnap's—so the dispute not merely terminological.

This attack on Putnam's position assumes elitism. If elitism is false, this sort of criticism of Putnam's nonsubstantivism will not succeed.

None of this would be news to Putnam. He responded to objections like that just mentioned by appealing to his doctrine of 'conceptual relativity', the view that 'the logical primitives themselves, and in particular the notion of object and existence, have a multitude of different uses rather than one absolute 'meaning";[16] had he used the word, Putnam would no doubt have added that none of these uses is elite.

I do not mean to imply that if elitism is false then Putnam's nonsubstantivism is correct—for there may be other problems with Putnam's position. The point, more

[16]From the first lecture of Putnam (1988).

modestly, is that *one particular line of criticism* to Putnam's nonsubstantivism is blocked is elitism is false.

5.3.4 Elitism and the debate between pragmatists and metaphysical realists

The debate in the 1980s between the 'pragmatists' Putnam and Rorty and the 'metaphysical realists' was complex: there were several different arguments going on at once. One of these arguments was about elitism. Rorty and Putnam defended egalitarianism, while some of their opponents were elitists.

We've already seen that Putnam rejected the idea that there is an elite quantifier—this was what his 'conceptual relativity' was about. Putnam was also rather dismissive of Lewis's idea that some properties are more natural than others, calling it 'spooky' and 'medieval sounding'.[17]

Rorty too opposed elitism. He made perhaps his clearest statement of his egalitarianism in his paper 'Method, Social Science, Social Hope', where he made fun of the idea that scientists are in the process of discovering 'Nature's Own Vocabulary'. I take it that by 'Nature's Own Vocabulary', he meant what I mean by 'elite vocabulary'.[18]

5.4 What elitism is

Roughly, then, elitism is the claim that one of the goals of inquiry is to construct theories that use elite vocabulary. Some notes of clarification are in order.

First, the goals of inquiry we're talking about are not the idiosyncratic goals

[17] See Lewis (1984).

[18] Rorty (1981). See also Rorty (1994a) and Rorty (1994b).

of particular researchers (getting tenure, impressing your husband, establishing a reputation as an intellectual maverick, or whatever). We're talking about the goals that inquirers have *as such*.

Second, the sort of inquiry that's at issue here is what's often called 'pure inquiry'. The chief characteristic of pure inquiry is that is not motivated by any particular practical goal—curing a disease, improving a building material, increasing the energy efficiency of a factory, or whatever. Instead, pure inquiry is motivated by curiosity.

Third, some properties of theories are valued only because they are taken to indicate other desirable properties. For example, proponents of Occam's razor think that theories with small ontologies are more likely to be true, other things being equal. Contrast this with the attitude of those who have a 'taste for desert landscapes'—people who think that having a small ontology is, all on its own, a desirable characteristic of a theory. Elitists think that, all on its own, containing elite vocabulary is a desirable characteristic of a theory—that this is not merely an indicator of some other desirable feature.

Here, then, is a more careful description of elitism. Elitists distinguish elite vocabulary from plebeian vocabulary, and claim that pure inquirers, as such, should seek to construct theories using elite vocabulary because this is a non-instrumentally desirable feature of such theories.

Having characterised this *full-blown* form of elitism, I should mention that there are some more modest versions—what I'll call *'partial'* forms of elitism.

We can imagine a metaphysician who accepts the distinction between elite and plebeian predicates, but rejects the distinction between elite and plebeian connectives. She might say, 'I can understand the idea that 'electron' is elite and 'dove' is not—but it doesn't make sense to worry about whether '$\wedge$' is an elite connective, or whether '$\rightarrow$' is elite'. This metaphysician is a partial elitist—she thinks that the elite/plebeian distinction applies only to words of certain semantic categories.

Why might someone be attracted to partial elitism? There are at least two reasons. First, and most obviously, the elite/plebeian distinction *feels* better when it is applied to some grammatical categories, and it feels worse when applied to other categories. Most people find the distinction between elite and plebeian predicates initially plausible—even if they ultimately reject it. The idea that there is an elite quantifier also has a certain appeal. The elite/plebeian distinction is much less attractive when applied to sentential connectives: most people are inclined to think that there is something wrong with the question 'Which of the truth-functional connectives is elite?'.

There's another, more theoretical reason for preferring a partial version of elitism. Some people will say that the distinction between elite and plebeian predicates is parasitic on a distinction between two sorts of property: perhaps Lewis's distinction between the perfectly natural properties and the less than perfectly natural properties. On this view, the elite/plebeian distinction won't apply to non-referring expressions—like perhaps sentential connectives.

My chief concern here is the full-blown form of elitism—though I will briefly come back to partial versions in my final section.

5.5 The epistemology of elitism

Enough with the preliminaries. Now I'm getting to the main part of the paper—I'm going to start explaining why I think that the elitist project is bound to fail. In this section, I'll discuss the epistemology of elitism. Granting for the moment that there is a distinction to be drawn between elite and plebeian words, how do we figure out which the elite ones are? What we need is a criterion for identifying elite expressions.

I only know of one reasonable proposal here, due to Sider.[19] In this section I'll present Sider's criterion.

Let's start with a simple question. Why is it so often said that terms from fundamental physics (like 'mass' perhaps) are elite, while aesthetic terms like 'delicious' are not? The obvious answer is that 'mass' occurs in a well-confirmed theory, whereas 'delicious' does not.

So here's a first stab at a criterion for identifying the elite terms:

> We have good reason to think a term elite if and only if it occurs in an epistemically virtuous theory.

I'm not going to say much about what 'epistemic virtue' is here—since elitists are free to disagree about this.

The proposal won't work. To see why not, consider the fact that some epistemically virtuous theories in economics might contain quantifiers that are restricted in some way: restricted to the things that are relevant to economics. The quantifiers in economic theory need not range over stars, for example. It doesn't follow that there is an elite quantifier which doesn't range over stars. The same goes for other disciplines which are limited in scope. Set theorists often use quantifiers that range over only sets. Even if set theory is epistemically virtuous, it doesn't follow that there is an elite quantifier that ranges over only sets.

To avoid this sort of problem, Sider follows Quine in talking not about theories (plural) but about our 'total theory'. The point is that individual theories might be limited in scope, so we should focus our attention on our best *overall* account of the nature of our world.

---

[19] See section 2.3 of Sider (2011).

So here's a revised suggestion:

> We have good reason to think a term elite if and only if it occurs in our most epistemically virtuous total theory.

For example, it's plausible that 'mass' will occur in our best total theory, but aesthetic terms won't.

But the criterion still isn't right. 'Our most epistemically virtuous theory' is a definite description, so it carries an implication or presupposition of uniqueness. And of course there might be a tie for the title 'most epistemically virtuous theory'.

For example, we can imagine an eighteenth century physicist who can't choose between a Newtonian version of physical theory and a Leibnizian one. The Newtonian theory uses a predicate 'x has position y', which relates an object to points in absolute space which it occupies. The Leibnizian physical theory uses the predicate 'x and y are distance z apart' instead. If the physicist is an elitist, what should he say about which of these predicates is elite? I think it's clear that, as long as the physicist is agnostic between the two theories, he should also be agnostic about which of the two theories uses elite vocabulary.

So we should correct our criterion like this:

> *The Indispensability Criterion*
>
> We have good reason to believe that an expression is elite just in case either it or a synonym occurs in all of our most epistemically virtuous total theories.

In calling this the 'indispensability criterion', I'm following Sider. Here's the idea. To dispense with an expression is to exhibit a maximally epistemically virtuous theory which doesn't contain the expression or a synonym. So we can rephrase our criterion like this: we have good reason to believe that an expression is elite just in case the expression is indispensable.

This is the only reasonable criterion that I know of for distinguishing the elite expressions from the rest. I'm going to be arguing that it doesn't enable us to identify any expressions as elite—basically because there are no expressions that are indispensable. My slogan is: if you think it's indispensable, you haven't tried hard enough. If I'm right about this, then if elitism is true, we can never know which expressions are elite. It will follow that the only viable form of elitism is the sceptical kind, and that the elitist project cannot succeed.

What remains is for me to give go through some examples of terms which look as though they might be elite, and argue in each case that the term is dispensable. Since this is a paper and not a book, I can't talk over all of my examples in great detail. So I'll discuss one example in some depth before briefly listing some other examples. I'll start by looking at the existential quantifier; this is a particularly important case, because as I've been saying many elitists put a lot of weight on the idea that there is an elite existential quantifier.

5.6 Dispensing with the existential quantifier

In this section, I'll argue that no existential quantifier is indispensable. The argument was inspired by a paper by John Burgess;[20] but of course I deserve the blame for its faults.

I'm going to dramatize the argument by imagining someone who believes what I'm going to call 'ontological nihilism'.[21] In everyday conversation, we talk as though the world were populated by *objects*—objects that have properties and bear relations to one another. The ontological nihilist thinks that this sort of talk is plebeian. She would like to find a different way of describing reality; she would like to find a language

[20]Burgess (2005).

[21]I've taken the term from Turner (2010), though Turner uses it in a different sense.

that is sufficiently expressively rich, but which doesn't contain proper names, or a first-order quantifier.

There'll be no problem in simple cases. Suppose for example that I say, 'There exists a rabbit!'. The ontological nihilist may say simply 'rabbit', meaning by this that the property of rabbithood is realized. Similarly, where I would say, 'There exists a brown rabbit!', she can say 'brown-and-also-rabbit'. And where I would say 'There exists something that is not a rabbit', she can say 'non-rabbit!'.

The strategy here is simple enough. She makes an assertion by uttering a predicate, and when she does so she means that the corresponding property is instantiated. Since she has only a limited lexicon, she needs to compound predicates much of the time. This is easy enough in simple cases, as we have seen, but she should prefer a systematic way of compounding predicates, and she would like to be sure that the resulting language is expressively rich enough. (How, for example, can she deal with 'there exists a rabbit, and something else which is brown'?).

Fortunately for her, Quine has already carried out this task, in his paper 'Variables Explained Away'.[22] He added to familiar, first-order predicate logic a family of 'predicate functors', which are used to compound simple predicates. I will now explain how Quine defined his predicate functors, before returning to my discussion of the ontological nihilist

The simplest of Quine's predicate functors is '$\sim$'; which 'negates' a predicate—so for example '$\sim$ *Dog*' is a predicate satisfied by all and only the non-dogs. More generally:

$[\sim Fx_1 \ldots x_n]$ is equivalent to $[\neg Fx_1 \ldots x_n]$

[22]Quine (1960).

'&' 'conjoins' predicates, like so:

$[(F\&G)x_1 \ldots x_m y_1 \ldots y_n]$ is equivalent to $[Fx_1 \ldots x_m \wedge Gy_1 \ldots y_n]$

For example, $(Dog\&Cat)$ is satisfied by x and y just in case x is a dog and y is a cat. 'Δ' is the 'derelativisation' predicate, defined like so:

$[\Delta F x_1 \ldots x_{n-1}]$ is equivalent to $[\exists x_n F x_1 \ldots x_n]$

For example, if '$Eats$' is a two-place predicate satisfied by x and y just in case x eats y, '$\Delta Eats$' is a 1-place predicate satisfied by all and only those things which eat something.

You might think that it would be useful to have a 'conversion' functor ('ϕ', say) which 'switches the argument places' of a binary predicate. So for example, x and y would satisfy '$\phi ParentOf$' just in case x is the child of y. As it turns out, to deal with predicates of adicity greater than two, it's better to have two different functors that behave like this, defined as follows:

$[\phi F x_1 \ldots x_n]$ is equivalent to $[Fx_1 \ldots x_{n-2} x_n x_{n-1}]$

$[\Phi F x_1 \ldots x_n]$ is equivalent to $[Fx_n x_1 \ldots x_{n-1}]$

Finally, there's a 'reflection' operator which decreases the adicity of its predicate:

$[\rho F x_1 \ldots x_n]$ is equivalent to $[Fx_1 \ldots x_n x_n]$.

For example, x satisfies $\rho Kills$ just in case x commits suicide.

This completes the list of functors.

An interesting feature of Quine's functors is that they enable one to construct 0-place predicates. I've already said that the derelativisation of the two-place predicate '$Eats$' is a 1-place predicate '$\Delta Eats$' which means *eats something*. Taking this one step further, '$\Delta\Delta Eats$' is a 0-place predicate which means *something eats something*.

With a bit of practice, it's straightforward to take sentences of first-order predicate logic and turn them into 0-place predicates, built out of atomic predicates and Quine's predicate-functors. Here's an example:

$\exists x \exists y(Rabbit(x) \wedge Brown(y) \wedge \neg x = y)$

$\exists x \exists y((Rabbit \& Brown)(x, y) \wedge \sim\!= (x, y))$

$\exists x \exists y((Rabbit \& Brown) \& \sim\!=)(x, y, x, y)$

$\exists x \exists y \phi((Rabbit \& Brown) \& \sim\!=)(x, y, y, x)$

$\exists x \exists y \Phi \phi((Rabbit \& Brown) \& \sim\!=)(y, y, x, x)$

$\exists x \exists y \rho \Phi \phi((Rabbit \& Brown) \& \sim\!=)(y, y, x)$

$\exists x \exists y \Phi \rho \Phi \phi((Rabbit \& Brown) \& \sim\!=)(y, x, y)$

$\exists x \exists y \Phi \Phi \rho \Phi \phi((Rabbit \& Brown) \& \sim\!=)(x, y, y)$

$\exists x \exists y \rho \Phi \Phi \rho \Phi \phi((Rabbit \& Brown) \& \sim\!=)(x, y)$

$\Delta \Delta \rho \Phi \Phi \rho \Phi \phi((Rabbit \& Brown) \& \sim\!=)$

The crucial result is that *any* name-free sentence of first order predicate logic is equivalent to some complex predicate, built up using only simple predicates and Quine's predicate-functors.[23]

Quine introduced his predicate-functors because he wanted to show that variables are in principle dispensable: we don't need to use them. In principle, we can avoid using variables by speaking 'Quinese'—the language whose sentences are just 0-place predicates constructed from atomic predicates using Quine's predicate functors.

[23]Perhaps I should explain this result slightly more carefully. It's fairly easy to see (based on the definitions of the predicate-functors given above) how to extend the standard definition of truth-at-a-model for standard first-order predicate logic to a definition of truth-at-a-model for the language that results from extending first-order predicate logic by adding Quine's predicate functors. Quine proved that for every name free sentence S of first-order predicate logic, there is a 0-place predicate P composed of just simple predicates and the predicate functors, such that S and P are true at precisely the same models.

As an aside, it's worth pointing out that we can extend Quinese so that sentences containing names can be translated into it: we need only add an appropriate sentence functor for each name. For example, 'Rabbit(Peter)' will become 'PeterRabbit', which is to be understood as meaning something like *the property Rabbit is instantiated Peterly*.

Back to the ontological nihilist. It looks as though Quinese is just what the ontological nihilist wanted: apparently, by speaking Quinese she can avoid using proper names or an existential quantifier, and Quine has shown that the resulting language is sufficiently expressively rich.

But there's a complication. Here again are Quine's definitions of the predicate functors:

$[\sim Fx_1 \dots x_n]$ is equivalent to $[\neg Fx_1 \dots x_n]$

$[(F\&G)x_1 \dots x_m y_1 \dots y_n]$ is equivalent to $[Fx_1 \dots x_m \wedge Gy_1 \dots y_n]$

$[\Delta Fx_1 \dots x_{n-1}]$ is equivalent to $[\exists Fx_1 \dots x_n]$

$[\phi Fx_1 \dots x_n]$ is equivalent to $[Fx_1 \dots x_{n-2}x_n x_{n-1}]$

$[\Phi Fx_1 \dots x_n]$ is equivalent to $[Fx_n x_1 \dots x_n]$

$[\rho Fx_1 \dots x_n]$ is equivalent to $[Fx_1 \dots x_n x_n]$.

Quine stipulated that the formulae in the left-hand column of this list are equivalent to the corresponding formulae on the right. The ontological nihilist will want to resist this claim. She wants to say that Quinese sentences are elite, while sentences containing quantifiers are not: so in at least one respect the two sorts of sentence are not 'equivalent'.

The nihilist will regard this list as describing a method for producing *inferior substitutes* for Quinese sentences, in the plebeian language of first-order predicate logic. Having reinterpreted Quinese in this way, the ontological nihilist can use the new language without reneging on her commitment to avoid the plebeian expressions of the 'ontological realists'.

My point is this. Most elitists think that some existential quantifier is elite—and they think that investigating the domain of this quantifier (the class of things which 'really exist') should be one of the goals of metaphysics. If I'm right, this is not justified. For all we know, there is no elite existential quantifier. The same thing works backwards. We can dispense with existential quantifiers by translating our theories into Quinese. In the same way, given a Quinese theory, we can dispense with the predicate functors by translating into the language of predicate logic. So both quantifiers, and the predicate functors, are dispensable. In the next couple of sections, I'll address a pair of objections to this claim.

5.7 First objection: Jason Turner

In his paper 'Ontological Nihilism',[24] Jason Turner argues that the derelativisation functor, 'Δ', is an existential quantifier. He agrees of course that it's a rather unfamiliar sort of quantifier, because it doesn't bind variables, but he thinks that it's a quantifier even so.

He has an argument for this conclusion, which I'll discuss in a moment. But even in the absence of such an argument, it's easy to see what he's getting at. Compare these two sentences:

$\exists x \exists y Eats(x, y)$

$\Delta\Delta Eats$

The 'Δ' in the latter sentence seems to be doing very much the same job as the '∃' in the former. The '∃' binds variables, while 'Δ' doesn't, but otherwise they look to be functioning rather similarly. So there's something intuitive about Turner's idea that the delta *just is* an existential quantifier.

[24] Turner (2010).

Turner's claim threatens my argument. I just said that we can dispense with existential quantifiers by using predicate functors. But if Turner is right, one of the predicate functors just is an existential quantifier. So let's take a look at Turner's argument.

5.7.1 Turner's Argument

The argument is complicated, so I'll break it up into three steps.

Step One: Variable-Binding and 'Quantification Proper'

Turner thinks that '$\exists$' does two jobs at once. As he puts it, the '$\exists$' symbol 'manages variable-binding, and it says something about how many values of its bound variable satisfy the postfixed formula'. He thinks it clarifies things to separate the two jobs, to separate 'variable binding' from 'quantification proper'. He goes on:

> This is what lambda-abstraction languages do. They have a predicate-forming operator, 'λ' that combines with a variable and an open expression to make a predicate: where 'ϕ' is an open expression, '$\lambda x \phi$' means 'is an x such that ϕ'. They also have expressions '$\exists_p$' and '$\forall_p$' that mean 'there is something that' and 'everything is such that' respectively.

For example, where in normal predicate logic we would write:

$$\exists x Rabbit(x)$$

Using Turner's notation we would write:

$$\exists_p \lambda x Rabbit(x)$$

By having one lambda within the scope of another, we can use this notation to create expressions for polyadic predicates. For example, here is an expression for the predicate x is y's brother:

$$\lambda y \lambda x (Sibling(x, y) \wedge Male(y))$$

Turner calls '$\exists_p$' a 'quantifier proper' (as opposed to '$\exists$', which is a quantifier-and-variable-binder-rolled-into-one). He claims, reasonably enough, that when one makes a statement whose leftmost symbol is '$\exists_p$', one has asserted *the existence of an object*. So the ontological nihilist will say that '$\exists_p$' is plebeian.

It will be helpful to have a new name for the new notation. I stipulate that 'the Lambda Language' is a modified version of first-order predicate logic in which lambdas and '$\exists_p$' are used in place of '$\exists$'. For simplicity I will suppose that there is no universal quantifier in the lambda language: neither '$\forall_p$' nor '$\forall$' is included.

Step Two: Lambda-Terms and Predicate-Functors

As I said, the following is a predicate meaning *is a brother of*:

$$\lambda y \lambda x(Sibling(x, y) \wedge Male(y))$$

So is this:

$$\rho(Sibling \& Male)$$

So these two predicates seem to mean the same thing. One uses lambdas and variables, the other uses predicate-functors; the effect is the same. Quine's achievement is to have shown that enough predicate-functors can do the same job as variables and lambdas. That's how Turner sees it, anyway. More generally, Turner thinks that if you take a Quinese predicate and excise the predicate-functors 'ρ', 'Φ', 'ϕ', '&', and '$\sim$' by using lambdas and variables instead, the result is a formula synonymous with the one you started with.

So for example, these are the same in meaning:[25]

$\rho\Phi\Phi\rho\Phi\phi((Rabbit\&Brown)\& \sim)$

$\lambda y\lambda x(Rabbit(x) \wedge Brown(y) \wedge \neg x = y)$

Turner introduces another language: I'll call it 'Lambda-Quinese'. Lambda-Quinese does contain 'Δ', but it doesn't contain the functors 'ρ', 'Φ', 'ϕ', '&', and '$\sim$'. Instead it contains lambdas and variables. You translate from Quinese to Lambda-Quinese in the manner I have just described.

Step Three: Comparing Lambda-Quinese and the Lambda Language

We've now created a modification of the language of predicate logic (the 'Lambda Language') and a modified version of Quinese ('Lambda-Quinese'). Now when you compare sentences from the two languages, they look very similar. To return to my example, consider the sentence:

$\exists x\exists y(rabbit(x) \wedge Brown(y) \wedge \neg x = y)$

This translates into the Lambda Language like so:

$\exists_p\exists_p\lambda y\lambda x(Rabbit(x) \wedge Brown(y) \wedge \neg x = y)$

---

[25]We need a convention about the order of the initial lambdas when constructing lambda terms. For example, why have I translated $\rho\Phi\Phi\rho\Phi\phi((Rabbit\&Brown)\& \sim)$ with:

(1) $\lambda y\lambda x(Rabbit(x) \wedge Brown(y) \wedge \neg x = y)$

rather than:

(2) $\lambda x\lambda y(Rabbit(x) \wedge Brown(y) \wedge \neg x = y)$?

My convention will be this. To translate a Quinese predicate, you begin by taking the corresponding predicate in the language that you get by adding Quine's predicate functors to standard predicate logic:

$\rho\Phi\Phi\rho\Phi\phi((Rabbit\&Brown)\& \sim)xy$

Then you ensure that the variables in the final translation are in the *opposite* order. So (1) rather than (2) is regarded as the correct translation.

In Quinese, the corresponding sentence is:

$$\Delta\Delta\rho\Phi\Phi\rho\Phi\phi((Rabbit\&Brown)\& \sim=)$$

In Lambda-Quinese, this becomes:

$$\Delta\Delta\lambda y\lambda x(Rabbit(x) \wedge Brown(y) \wedge \neq x = y)$$

Now if you look at the Lambda-Quinese sentence and compare it to the sentence from the Lambda-Language, they're almost the same. The only difference is that in Lambda-Quinese, 'Δ' replaces '$\exists_p$'. This works generally: given any sentence of the Lambda-Language, the corresponding sentence of Lambda-Quinese is obtained by replacing each 'Δ' with '$\exists_p$'.[26]

For Turner, this is enough to show that 'Δ' and '$\exists_p$' mean the same thing, by the following general principle:

> *Turner's translation principle*
>
> Suppose L_1 and L_2 are languages that are exactly alike except that, where L_1 has an expression α, L_2 has a different expression, β. If ϕ is a sentence in L_1 that uses α, we write it as 'ϕ_α', and 'ϕ_β' will be the expression of L_2 that is just like 'ϕ_α' except that β is replaced everywhere for α. ... If every term (other than α and β) is interpreted the same way in L_1 as it is in L_2, and if the speakers of L_1 utter ϕ_α in all and only the circumstances in which speakers of L_2 utter ϕ_β, then α and β have the same interpretation also.[27]

---

[26]The order in which the variables are bound by lambdas in the Lambda-Quinese sentence matters a great deal here. That's why I introduced the convention in footnote 25.

[27]Turner spends some time in his paper clarifying and defending this principle—I'm omitting details from his discussion that don't affect my argument. See section 4.1.2 of Turner (2010).

In our case, L_1 is the Lambda Language, and L_2 is Lambda-Quinese. We can assume that the two languages contain the very same stock of simple predicates. We've constructed them so that they have the same connectives and variables, and 'λ' means the same thing in both cases (see *Step Two*). Applying Turner's translation principle with α as '$\exists_p$', and β as Δ, we conclude that 'Δ' and '$\exists_p$' are synonymous—which is Turner's desired conclusion. As we saw in *Step One*, '$\exists_p$' is a 'quantifier proper'. Hence, so is 'Δ'.[28]

5.7.2 My response to Turner's argument

Before getting into my response, I'd like to pause to say something about the dialectic, which has gotten a bit confusing. Back in section six, I introduced ontological nihilism—a position according to which Quine's predicate functors are elite while existential quantifiers (and proper names) are plebeian. My own view is that, assuming elitism, we have no way to choose between this ontological nihilist position and an alternative position according to which there is an elite existential quantifier. If I'm right about this, elitism has a sceptical consequence: we *don't know* whether there is an elite existential quantifier, or whether Quine's predicate functors are elite. In this section, I'm discussing Jason Turner's claim that the derelativisation functor just is an existential quantifier. If Turner is correct about this, then ontological nihilism is untenable—and this would undermine my claim that elitism has this sceptical consequence. This leaves me with the task of defending ontological nihilism from Turner's criticism. This is a strange position for me to be in, since as an egalitarian I reject ontological nihilism. Philosophy makes strange bedfellows.

---

[28] I've actually modified the argument in one small respect. Turner's version of the argument involves an extra symbol: δ. I've changed the argument to remove the need for this extra symbol. I find the plethora of new symbols involved in the argument already rather confusing. The discussion is not affected in any important way by this change. Sceptical readers should compare my presentation of the argument with section 6.3.1 of Turner (2010).

I suggest that the ontological nihilist should begin by conceding (for the sake of argument at any rate) that 'Δ' is an existential quantifier. I will suggest a new, improved version of Quinese ('New Quinese', I'll call it) in which 'Δ' is replaced with a different symbol '$\cap$'. I'll ensure (i) that it is not an existential quantifier, and (ii) that New Quinese is not expressively impoverished.

As before, I'll introduce the functor by showing how to produce substitutes for sentences containing the functor in the language of first-order predicate logic:

$[(F \cap G)x_1 \ldots x_{m-1}y_1 \ldots y_{n-1}]$

translates to

$[\exists z(Fx_1 \ldots c_{m-1}z \wedge Gy_1 \ldots y_{n-1}z]$

New Quinese is the language that consists of predicates composed using only the functors 'ρ', 'Φ', 'ϕ', '&', '$\sim$', and '$\cap$'.

In practice, the two Quineses don't differ much. Suppose F and G are 1-place predicates. Where in Old Quinese one says 'ΔF', in New Quinese one says '$(F \cap F)$'.[29] Where in New Quinese one says '$(F \cap G)$', in Old Quinese one says '$\Delta\rho(F\&G)$'.[30]

Now, I claim that New Quinese has no existential quantifier in it—and so, as I said before, the existential quantifier is dispensable.

Jonathan Schaffer, on behalf of the critic of ontological nihilism, has pointed out to me a clever response to this. The critic can begin by introducing a new symbol

---

[29]This generalizes immediately to the case in which F is polyadic, or not primitive.

[30]The general case is more difficult. Suppose F is a primitive m-place predicate, and G is primitive n-place predicate. Then where in New Quinese one says:

$(F \cap G)$

in Old Quinese one says:

$\Delta\rho(\Phi^{m+n-1}\phi)^{m-1}(G\&F)$

where the superscripts represent repeated applications of the operators in the obvious way. This generalizes immediately to the case in which F and G are non-primitive predicates.

'$\sqcap$' into the lambda-language with the following stipulation:

> $[\alpha \sqcap \beta]$ is analytically equivalent to $[\exists_p \lambda z(\alpha(z) \wedge \beta(z))]$, where z does not occur free in α or β.

He can then, using the techniques explained in section 7.1, argue that '$\cap$' and '$\sqcap$' are synonyms.

Now the nihilist is currently defending the view that '$\cap$' is elite. This implies, given that '$\cap$' and '$\sqcap$' are synonyms, that '$\sqcap$' is elite. But this seems to imply that '$\exists_p$' is elite (since '$\exists_p$' is needed for the definition of '$\sqcap$'). So once again, it seems that the ontological nihilist is unable to sustain his claim that there is no elite existential quantifier. I'll call this argument 'the countercounterattack'.

There's something very odd about the countercounterattack. The nihilist's critic is now defending the view that these two formulae are synonymous, despite the difference in their logical forms:

$F \cap G$

$\exists_p \lambda x(Fz \wedge Gx)$

This is strange because elitists are usually hostile to the idea that sentences with different logical forms are synonymous, as we saw back in section 3.3. This raises the suspicion that the countercounterattack uses the translation principle in a way that should not be acceptable from an elitist point of view.

We can see that this suspicion is right by considering this argument, whose conclusion is that if any monadic predicate F is elite then so is every other monadic predicate G:

> Define F^* as $(F \wedge G) \vee (\neg F \wedge \neg G)$, and F^{**} as $(F^* \wedge G^*) \vee (\neg F^* \wedge \neg G^*)$. We can use Turner's translation principle to show that F and F^{**} are synonyms. Since F is elite, so is F^{**}. But since F^{**} is defined using G, G must then be elite too.

Clearly, the elitist will have to say that there is something wrong with this argument—but what? I see three options ...

Option 1: It might be said that it is illegitimate to apply Turner's translation principle to terms which have analytic definitions. The thought here would be that the translation principle has to be restricted to words that don't have analytic definitions, because the unrestricted version of the principle is too strong since it implies that there are synonymous sentences with different logical forms. In particular, it might be said, the translation principle should not be used to argue that F^{**} and F are synonymous. If this is right, then of course it is also illegitimate to use the translation principle to argue that '$\cap$' and '$\sqcap$' are synonymous.

Option 2: The second option would be to deny that from 'F is elite' and 'F is defined using G', it follows that G is elite. If this is right, then even if it can be shown that '$\sqcap$' is elite, it won't follow that '$\exists_p$' is elite—and so again the countercounterattack fails.

Option 3: One might try to block the argument for the conclusion that F and F^{**} are synonymous by arguing that they are not intersubstitutable in some intensional context. For example, it might be claimed that F^{**}-facts are grounded in G-facts, but F-facts are not, and so F^{**} and F are not substitutable when embedded under whatever operator is used to make claims about grounding. If this is right, then the nihilist can say that $\sqcap$-facts are grounded in $\exists_p$-facts, but $\cap$-facts are not, and this undermines the argument for the conclusion that $\sqcap$ and $\cap$ are synonyms.

If I am right, then if the countercounterattack is sound, so is my argument for the conclusion that if one monadic predicate is elite, so are all the others. So the nihilist's position is secure against attacks by other naïve elitists.

5.8 Second objection: Shamik Dasgupta

In the abstract to his paper 'Individuals, An Essay in Revisionary Metaphysics',[31] Shamik Dasgupta summarises his position like this:

> We naturally think of the material world as being populated by a large number of individuals. These are things, such as my laptop and the particles that compose it, that we describe as being propertied and related in various ways when we describe the material world around us. In this paper I argue that, fundamentally speaking at least, there are no such things as material individuals.

Dasgupta goes on to say that to describe the fundamental facts about the physical world, it is Quinese (rather than predicate logic) that we should use.

Dasgupta's argument concerns what is fundamental, rather than what is elite, so his argument is not directly relevant. But it's plausible that words needed to describe fundamental facts are elite, so his argument is pertinent in an indirect way.

I've been saying that we have no idea whether it is the quantificational apparatus of predicate logic, or the functors of Quinese which are elite. If Dasgupta is right, however, the conclusion can be avoided. We can be confident that it is the predicate functors which are elite, and not the quantifiers and variables of predicate logic.

In section 8.1 I'll explain Dasgupta's argument; I'll criticise it in section 8.2.

5.8.1 Dasgupta's argument

Dasgupta begins by attacking a view which he calls 'individualism'. According to individualists, 'the most basic, irreducible facts about our world include facts about what individuals there are and how they are propertied and related to one another'.

[31] Dasgupta (2009).

On this view, the fundamental facts about the material world are what he calls 'individualistic facts'—facts like:

> a is F, b is G, a bears R to b,

Dasgupta's case against individualism, which will be described only in outline here, begins with an analogy. NGT is a fragment of classical mechanics which consists of Newton's laws of motion together with his law of gravity. We can distinguish two different versions of this theory:

> $\mathrm{NGT_A}$: NGT combined with an absolute theory of space (i.e. a theory according to which there are facts about the absolute positions and velocities of particles).[32]
>
> $\mathrm{NGT_R}$: NGT combined with a relational theory of space (i.e. a theory according to which there are no facts about the absolute positions and velocities of particles).

It is widely felt that $\mathrm{NGT_R}$ is superior to $\mathrm{NGT_A}$, for the following reason. $\mathrm{NGT_A}$ posits facts about absolute position and velocity, but the laws of NGT are (in a sense that can be made precise) *insensitive* to these facts. Facts about *relative* position are important in NGT, because they help to determine the forces that act on the particles. And the laws specify the way in which the *acceleration* of each particle is determined by its mass and the forces that act on it. But, to put it loosely again, the laws don't constrain the absolute positions and velocities of the particles at all, except by constraining their accelerations and relative positions. Now there seems to be something objectionable about positing facts about absolute position and velocity,

[32]Dasgupta's discussion can be reformulated so as to avoid this reification of facts—but to keep things simple in following Dasgupta by indulging in this reification.

when these facts are redundant: NGT_A seems to be overly complicated, inelegant and unlovable for this reason.[33]

Dasgupta thinks that individualism is unlovable for a similar reason: individualistic facts are redundant and so we should avoid positing them if we can. Suppose we have some particles moving around according to the physical laws; facts about the relative positions of these particles might be important, as might facts about their charges, or their masses, or facts about the structure of space, or facts about the wave-function of the universe—and so on. But, barring some huge surprise in physics, it doesn't matter at all which particle is Alfredo, or which is Benedetta. The laws of physics are insensitive to facts about which individual is which; individualistic facts are redundant. And so we do better to omit such facts from physical theory.

Having rejected individualism, Dasgupta goes on to look at an alternative, which I will call 'naïve generalism'. On this view, the fundamental facts are *quantificational*—facts of the form:

$$\exists x Fx;\ \exists y Gy;\ \exists x \exists y Rxy;\ \ldots$$

Dasgupta rejects naïve generalism with a quick argument:

> [Naïve generalism] is unacceptable. After all, we have been brought up to understand that quantifers range over a domain of individuals. So our natural understanding of the facts listed above is that they hold in virtue of facts about individuals, and it would therefore appear that we have made no progress.

Put another way, the objection is this. Dasgupta thinks that an existential fact, a fact of the form $\ulcorner \exists x \phi x \urcorner$ must be grounded in a 'witnessing fact', a fact of the form $\ulcorner \phi a \urcorner$.

---

[33] Dasgupta also makes an epistemic point: absolute velocities cannot be measured, if NGT is true—and so NGT_A posits facts which epistemically inaccessible to us. To save on space I don't discuss this idea.

It follows that existential facts cannot be fundamental. So Dasgupta rejects naïve generalism. In its place, he advocates an alternative form of generalism (Dasguptan generalism?) according to which the fundamental facts are such as to be properly described using Quinese.

5.8.2 A response to Dasgupta

I am not going to criticise Dasgupta's argument against individualism. For the record, I think it's a strong argument. I am going to do something much more modest: I'm going to criticise Dasgupta's argument against naïve generalism. Dasupta's main concern seems to be the 'generalism vs. individualism' issue, so perhaps it doesn't matter so much to him which form of generalism is true. (This would explain why his criticism of naïve generalism is terse). So perhaps Dasgupta would regard my criticism as somewhat peripheral.

As we've seen, in making his argument against naïve generalism, Dasgupta uses this premise:

> *The Existential Grounding Thesis*
>
> Every fact of the form $\ulcorner \exists x \phi x \urcorner$ is grounded in a fact of the form $\ulcorner \phi a \urcorner$.

I think that this premise is not well motivated, and so Dasgupta's argument is unconvincing.

In this passage, which I have already quoted, Dasgupta hints that he has an argument for the existential grounding thesis:

> [Naïve generalism] is unacceptable. After all, we have been brought up to understand that quantifers range over a domain of individuals. So our natural understanding of the facts listed above is that they hold in virtue of facts about individuals, and it would therefore appear that we have made no progress.

Apparently, Dasgupta is trying to derive the existential grounding thesis from the standard model-theoretic treatment of the quantifiers of predicate logic. Now it is true that the standard model-theory for predicate logic has this consequence:

> The statement '$\exists x F x$' is true at a model $\mathcal{M}$ just in case some element of the domain of $\mathcal{M}$ is also an element of the set that $\mathcal{M}$ assigns to 'F'.

But it doesn't follow that:

> If '$\exists x F x$' is true, then the fact that $\exists x F x$ is grounded in the fact that some individual satisfies F.

Analogously, you can't establish that facts expressed using second-order quantifiers are grounded in facts about the universe of sets, just by appeal to the fact that the standard model theory for second-order logic is set-theoretic. And you can't establish that modal facts are grounded in facts about possible worlds just by appeal to the standard Kripke model theory for modal languages.

But maybe I am being unfair to Dasgupta: perhaps he didn't intend to offer an argument for the existential grounding thesis, perhaps he just took it as an assumption, thinking it obviously true.

I agree that the existential grounding thesis is very plausible, but I don't think that it is legitimate for Dasgupta to appeal to this intuition at this stage in his argument. To see why not, think about *why* the existential grounding thesis is plausible. The reason, I think, is that we have in mind a sort of hierarchical picture of how quantified facts relate to one another—I'm going to call this the 'Fregean Hierarchy', though I wouldn't be surprised to learn that it actually predates Frege.

Here's the picture:

. . .

. . .

Third Floor: $\exists X Instantiated(X)$; $\exists X More(F, X)$

Second Floor: $\exists x Fx$; $\exists x \exists y Rxy$; $Instantiated(G)$

Ground Floor: Fa; Gb, Rab;

On the ground floor there are individualistic facts like Fa, Gb and Rab. On the second floor there are facts about first-order concepts, facts like $\exists x Fx$ and the fact that G is instantiated. Then on the higher floors there are facts about higher-order concepts.

Now when making his case against naïve generalism, Dasgupta assumes that every fact of the form $\ulcorner \exists x \phi x \urcorner$ is grounded in a fact of the form $\ulcorner \phi a \urcorner$—this is the existential grounding thesis. I think that the thesis is plausible because we have the Fregean hierarchy in mind when we think about quantification. Now the problem is that Dasgupta has just given us an argument for lopping off the ground floor of this hierarchy—he's given us an argument for rejecting the claim that individualistic facts are fundamental. Now if Dasgupta's argument is convincing, this should motivate us to reject or modify the Fregean hierarchy—and this undermines the motivation for the existential grounding thesis. Dasgupta is appealing to an intuition that he himself has just undermined!

5.9 How to dispense with some other terms

5.9.1 Dispensing with mathematical vocabulary

Many of the philosophers who have opposed the elitist project have worked on the philosophy of maths. This includes Carnap and Putnam, and more recently John

Burgess.[34] This is no coincidence: mathematics is problematic for elitists, as we'll see.

People say that 'mathematics is indispensable'. This slogan can be misleading: I accept the received view that mathematics as a whole is indispensable; nevertheless, I will argue that no *particular* mathematical expression is indispensable.

As is well known, every expression of standard mathematics can be defined *eliminatively* using the basic predicates of set theory, specifically 'x is a set' and 'x is an element of y'—expressions which are not usually given definitions. This already shows that all standard mathematical vocabulary is dispensable, except perhaps these two basic set-theoretic predicates. The vocabularies of number theory, analysis, geometry, topology and so on, are all dispensable.

For a simple and familiar example, let's consider the natural numbers. As is well known, these numbers can be identified with sets—for example in this way:

$$
\begin{array}{lll}
0 & := \varnothing & \\
1 & := \{0\} & = \{\varnothing\} \\
2 & := \{0, 1\} & = \{\varnothing, \{\varnothing\}\} \\
3 & := \{0, 1, 2\} & = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\} \\
\ldots & & \\
\ldots & &
\end{array}
$$

Having made this identification, one can explicitly define all the vocabulary of number theory in set-theoretic terms. For example, we can define '$+$' and '$\times$' set-theoretically. Having done this, one can regard the theorems of number theory as mere abbreviations

[34]See Carnap (1950); Putnam (1988) and Putnam (2004); and the essays in Burgess (2008), especially Burgess (2005).

of theorems of set theory. We have dispensed with the vocabulary of number theory. It is a great achievement of twentieth-century mathematicians to have shown in detail that this is correct, and to have similarly dispensed with the vocabulary of analysis, topology, geometry and the rest.

Now it turns out that set-theoretic vocabulary is dispensable too: we can avoid the predicates of set theory by using the theory of functions instead. The basic idea is that just as we can avoid number-theoretic vocabulary by 'replacing' numbers with sets, so we can avoid set-theoretic vocabulary by 'replacing' sets with functions. For example, we could 'replace' each set S with a function F_S such that:

For all $x \in S$, $F_S(x) = 0$

For all $x \notin S$, F_S is undefined at x.[35]

---

[35]I should suggest some axioms for an autonomous function theory. I'll write the axioms informally in English, but they can be formalized in first-order logic with the sole non-logical expression '$f : x \to y$', for 'f maps x to y'. I'll say that a function is 'small' when it is in the domain or range of another function.
Extensionality Axiom: If $\forall X, Y$, $F : X \to Y \leftrightarrow G : X \to Y$, then $F = G$.
Functionality Axiom: If $F : X \to Y_1$, and $F : X \to Y_2$, then $Y_1 = Y_2$.
Empty Function Axiom: There is an 'empty' function, e, such that for no X and Y is it the case that $e : X \to Y$; e is small.
Function Extending Axiom: For any small function f undefined at x where is x is small, and for any small y, there is a small function f' such that:

- $f'(z) = f(z)$ for any z in the domain of f.
- $f'(x) = y$
- f' is otherwise undefined

Union Axiom: Suppose f is a small function with no values except perhaps e; then there exists a small function $(\cup f)$ such that:

$(\cup f)(x) = y$ just in case for some f' such that $f(f') = e$, $f'(x) = y$.

Powerfunction Axiom: Suppose f is a small function, then there exists a small function $(\mathcal{P}f)$ such that $(\mathcal{P}f)(x) = e$ if x is a subfunction of f, otherwise $(\mathcal{P}f)$ is undefined.
Axiom of Infinity: For any small function f, let f^+ be a function such that:

- If $f : x \to y$, $f^+ : x \to y$.
- $f^+ : f \to f$
- f^+ is otherwise undefined.

There exists a small function ω such that for any small function f, if $f \subseteq \omega$ then $f^+ \subseteq \omega$.
Function Existence Axiom Schema: Suppose that $\phi(X, Y)$ is a formula whose free variables include X and Y. Then if for every X there is exactly one Y such that $\phi(X, Y)$, then there is a function F

So to repeat, it's a well-known fact that all mathematical vocabulary is dispensable with the help of set theory. And one can dispense with the language of set theory using the theory of functions. So even though any serious total theory is going to have to contain some mathematical vocabulary, no particular mathematical word is indispensable. To use Sider's 'structure' metaphor, elitists must conclude that the world has a mathematical structure which is, and always will be, hidden.

The example also has ramifications for the elitist idea that there is a privileged existential quantifier. If the elitists are right, it ought to be a serious question whether it is the sets that 'really' exist or the functions. But if this is a serious question, we'll never know the answer to it.

5.9.2 Dispensing with proper names

Here's another Quinean trick.[36] We can eliminate names from our language by replacing them with bespoke one-place predicates. For example, we can avoid using the name 'Socrates' so long as we have a one-place predicate 'socratize' (which is true of Socrates and nobody else). For example, instead of saying 'Socrates was wise', one can say 'The person who socratized was wise'.

---

such that for all X and Y, $F : X \to Y$ iff $\phi(X, Y)$.
Restriction Axiom: Given any small function f and any function G, there is a small function h such that for all X and Y, $h : X \to Y$ just in case $f : X \to Y$ and $G : X \to Y$.
Replacement Axiom: Given a small function f and any other functions A and B, there is a small function f' such that $f' : X \to Y$ just in case for some X' and Y', $A(X') = X$, $Y(Y') = Y$, and $f : X' \to Y'$.
ZFC set theory can be interpreted in this theory, in the manner suggested in the text. The theory can be interpreted in NBG set theory by understanding the '$f : x \to y$' notation in the obvious way, and taking the quantifiers to range over the 'pure functions', where a pure function is a function all of whose arguments and values are functions all of whose arguments are functions, all of whose ... and so on.

[36] Quine (1948).

5.9.3 Dispensing with higher-order quantifiers

Higher-order quantifiers can be eliminated using the vocabulary of set theory, or by repeating the predicate functor trick 'one level up' as it were.

5.9.4 Determinates and determinables

There are several ways of formalising statements about determinates and determinables. One way of doing it is to introduce a 1-place predicate for each determinate. To use mass as an example, one would introduce the predicates '$\mathrm{M}_{3\mathrm{kg}}$', '$\mathrm{M}_{1.7\mathrm{kg}}$', '$\mathrm{M}_{100\mathrm{g}}$'—and so on.

But all of these predicates are dispensable. Instead of using these predicates, one could extend one's ontology with a family of *sui generis* objects ('masses'). One could then formalise one's statements about mass using a two-place predicate 'x has mass y', and also one or more predicates to describe the relations between the masses.

There's another, more colourful way of doing things. The *sui generis* objects, masses, form a 1-dimensional space. One might suppose that to 'have a mass' is simply to have a location in this one-dimensional space. On this view, physical objects are multiply located: at any one time they are located in physical space, and also in 'mass space' (and perhaps also in 'charge space' etc.). When formalising this theory, one would make use of some terminology used to characterise the various spaces, and then a two-place predicate 'x occupies point y' to locate objects in the various spaces that they occupy. Thus, one can dispense with the 'x has mass y' predicate by using the 'x occupies point y' predicate.

So even if we need in our theory some vocabulary for talking about mass, no particular expression about masses is indispensable.

5.10 Some objections to the argument

5.10.1 Eliteness and intuition

I have been arguing that, if elitism is true at all, we cannot figure out which the elite expressions are. A critic might respond by saying that we can figure this out, by making more extensive appeals to intuition than I have been allowing so far. For example, the critic might say that we can reject the proposal that we dispense with the predicate 'x has mass y' by saying that objects are located in both mass space and physical space, by appealing to the intuition that individuals (unlike universals) cannot be multiply located.

I have no single response to this sort of criticism: I think that such arguments should be dealt with one-by-one. I anticipate that such appeals to intuition will not discredit the claims of this paper—but we must wait and see.

As a sort of miniature case study, I will consider the intuition that I just mentioned, according to which individuals cannot be multiply located. Contrast the two claims:

(1) No individual can have two locations within any one space.

(2) No individual can have two locations.

The defender of the multiple-location view must deny (2), of course—and perhaps this is a weakness of her view. However, she can respond by pointing to a closely related strength. While she must reject (2) she can accept the superficially similar (1). This is related to an attractive feature of her view: it allows her a unified explanation of the fact that physical objects can have only one location in physical space, and the fact that they can have only one determinate property for each determinable. So it's not clear that, on balance, the multiple-location theory is unintuitive.

5.10.2 Only mildly sceptical forms of elitism

I've been distinguishing naïve forms of elitism from sceptical ones. The sceptical elitists claim that we cannot figure out which the elite terms are; the naïve ones deny this claim. A critic might say that this is a rather crude way of thinking about it—after all, it seems that scepticism comes in degrees. At one extreme are those who claim that can never have any idea at all which expressions are elite; at the other extreme, I suppose, would be elitists who claim already to know exactly which the elite words are. There is a need, the critic might add, to investigate *only mildly sceptical* forms of elitism. Mildly sceptical elitists concede that we may never be able to draw up a definitive list of elite terms, but think it reasonable to hope that we will be able to make significant progress towards this goal—and so the elitist project is not totally moribund.

It's a fair point. So in this sub-section I will discuss a couple of mildly sceptical forms of elitism.

First Variant

Naturalness, to repeat, is said to come in degrees. At one extreme are the most natural properties—what Lewis called the 'perfectly natural properties'. The property of having mass 1kg might be one of these. The property of being a lion will be less than perfectly natural; the property of being a dove is still further from perfect naturalness; and all of these properties are more natural than the property of being either a dove or an aqueduct, or a prime number that isn't seventeen.

In his discussions of this subject, Lewis cautiously suggested the following principle, which is easily generalised to other parts of speech:[37]

> *The length of definition principle*
>
> A predicate F is more natural than a predicate G just in case the shortest definition of F in *perfectly natural terms* is shorter than the shortest definition of G in *perfectly natural terms*.

Here's an example. A proponent of the length of definition principle will say that if 'has mass 1kg' and 'has charge 1C' are perfectly natural predicates, then 'has mass 1kg and charge 1C' is *almost* perfectly natural, because it is so quick to define in perfectly natural terms. In contrast, it takes a *lot* of work to define 'dove' in perfectly natural terms, which is why 'dove' is an *un*natural predicate.

Now a mildly sceptical elitist might object to my argument in the following sort of way:

> Perhaps you're right that we don't know that some existential quantifier is elite, because one can dispense with quantifiers by using Quinese. However, since the existential quantifier is easy to define in Quinese, if Quinese expressions are elite, the existential quantifier is *almost elite*, by the length of definition principle. So you haven't refuted the following, more modest claim: we can be sure that that some existential quantifier is elite or *almost elite*. This is disappointing to be sure, but it hardly motivates your conviction that we should just give up on the elitist project altogether.

[37] See Lewis (1986), pg. 61.

My response to this is simple: even if we set aside general concerns about the standing of the concept of naturalness, the length of definition principle is false.[38] It is quite easy to see that there are very unnatural predicates which have short definitions in perfectly natural terms. Consider, for example, the predicate 'has a mass in kilograms whose integer part is a prime number'.[39]

Pending some dramatic revision of the length of definition principle, this form of mildly sceptical elitism should be rejected.

Second Variant

I've been arguing that if elitism is true, we have no way of identifying the elite expressions; however, my argument does not show that we are unable to identify some words as plebeian. For example, I have not argued that we don't know that 'grue' is plebeian. A critic might say:

> You may be right that we will never be able to draw up a list knowing that all and only the elite expressions appear on it. But perhaps we will be able to draw up a list of expressions and say, with justification, 'Some of these expressions are elite, and some are plebeian, but all of the elite expressions are on the list'. Perhaps both the predicate functors and the expressions of ordinary predicate logic will be on the list. This could be an important achievement, if the list is short enough.

Personally, I am pessimistic about this project: I doubt that such a list could ever be 'short enough', and I hope that the examples I've discussed in this paper justify this scepticism. But we must wait and see. Sceptics like me should continue to devise

---

[38]Thanks to Brian Weatherson for explaining this to me—using more detail than I am including here.

[39]I'm assuming here that some mass vocabulary is elite. If this is wrong, some other determinable can be used.

tricks for dispensing with supposedly elite words; naïve elitists should continue trying to show that this or that word is plebeian. My prediction is that our list will never be short.

5.11 Conclusion

To finish off, I'll summarise what I hope to have shown. I only know of one reasonable account of how we might figure out which the elite words are. According to this account, we have reason to believe that a term is elite only if it is indispensable from our most epistemically virtuous theories. I've argued by giving examples that *nothing* is indispensable. So I think we'll never have good reason to think that a term is elite. The elitist project, or at least the more ambitious forms of it, cannot succeed.

It is not only *full-blown* naïve elitism that is threatened by my argument, but also *partial* naïve elitism. Elitism is most plausible when applied to the existential quantifier, and to predicates. I have shown that no existential quantifier is indispensable. I have also shown how to dispense with various predicates that one might otherwise have taken to be elite: *set*, *number*, *is the mass of*, ...

If you think it's indispensable, you haven't tried hard enough. If you think it's elite, your belief is not justified.

Bibliography

Alston. *A Realist Conception of Truth*. Cornell University Press, 1996. 56

Armstrong. *Universals and Scientific Realism*. Cambridge University Press, Cambridge, 1978. 139

Austin and Frege. *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*. Blackwell, Oxford, 1974. 93

Ayer. *Language,Truth and Logic*. London, V. Gollancz, Ltd., 1936. 125

Azzouni. *Deflating Existential Consequence: A Case for Nominalism*. Oxford University Press, Oxford, 2006. 141

Bach. Truth, justification and the american way. *Pacific Journal Quarterly*, 73:16–30, 1992. 62

Banks. *Ernst Mach's World Elements: A Study in Natural Philosophy*. Kluwer Academic Publishers, Dordrecht, 2003. 21

Benacerraf. What numbers could not be. *Philosophical Review*, 74(1):47–73, 1965. 127

Bishop and Murphy, editors. *Stich and His Critics*. Wiley-Blackwell, 2009. 62

Boghossian. Analyticity reconsidered. *Noûs*, 30(3):360–391, 1996. 125

Bourbaki. *Théorie des Ensembles*, volume 1 of *Eléments de Mathématiques*. Hermann, Paris, 1 edition, 1939. 128

Burgess. Being explained away. *Harvard Review of Philosophy*, 13:41–56, 2005. 151, 171

Burgess. *Mathematics, Models and Modality*. Cambridge University Press, Cambridge, 2008. 171

Carnap. *Der Logische Aufbau der Welt*. Weltkreis, 1928. 46

Carnap. Empiricism, semantics and ontology. *Revue Internationale de Philosophie*, 4:20–40, 1950. 11, 171

Cauchy, Bradley, and Sandifer. *Cauchy's Cours d'Analyse*. Springer, 2009. 130

Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 1996. 42

Churchland. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78:67–90, 1981. 1

Clifford. On the nature of things-in-themselves. *Mind*, 3:57–67, 1878. 3, 21

Dasgupta. Individuals: An essay in revisionary metaphysics. *Philosophical Studies*, 145(1):35–67, 2009. 165

Dudley. *Mathematical Cranks*. The Mathematical Association of America, 1992. 125

Dworkin. Objectivity and truth: You'd better believe it. *Philosophy and Public Affairs*, 25(2):87–139, 1996. 143

Ebert and Shapiro. The good, the bad and the ugly. *Synthese*, 170(3):415–441, 2009. 109

Field. *Science Without Numbers*. Princeton University Press, 1980. 2, 10

Fine. The question of realism. *Philosophers' Imprint*, 1(2):1–30, 2001. 39, 141, 143

Fine. *The Limits of Abstraction*. Oxford University Press, 2002. 94, 95, 104, 110

Frege. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung ber den Begriff der Zahl*. Koebner, Breslau, 1884. 93

Friedman. Carnap's aufbau reconsidered. *Noûs*, 21(4):521–45, 1987. 46

Goldman. Review of stephen p. stich: The fragmentation of reason. *Philosophy and Phenomenological Research*, 51(1):189–193, 1991. 56

Goldman. *Knowledge in a Social World*. Oxford University Press, 1999. 66, 69, 71

Hale. Reals by abstraction. In Bob Hale and Crispin Wright, editors, *The Reason's Proper Study*, pages 399–420. Clarendon Press, 2001. 124

Hale and Wright. Implicit definition and the a priori. In Boghossian and Peacocke, editors, *New Essays on the A Priori*. Oxford University Press, 2000. 93, 125

Hale and Wright. Focus restored: Comments on john macfarlane. *Synthese*, 170(3): 457–482, 2009a. 111, 130

Hale and Wright. The metaontology of abstraction. In David Chalmers, David Manley, and Ryan Wasserman, editors, *Metametaphysics*, pages 178–212. Oxford University Press, Oxford, 2009b. 93, 113

Harman. Justification, truth, goals and pragmatism. *Philosophy and Phenomenological Research*, 51(1):195–99, 1991. 64

Hilbert. *Grundlagen der Geometrie*. Teubner, Leipzig, 1899. 129

James. The sentiment of rationality. In *The Will to Believe and Other Essays in Popular Philosophy*. Harvard University Press, Cambridge, MA, 1879a. 29

James. The spatial quale. *Journal of Speculative Philosophy*, 13(1):64–87, 1879b. 24

James. *The Principles of Psychology*. Henry Holt and Company, New York, 1890. 5, 24, 25, 29

James. *The Varieties of Religious Experience*. Longman, Green and Company, 1902. 27

James. A world of pure experience. *Journal of Philosophy, Psychology and Scientific Methods*, 1(21):561–70, 1904a. 4, 22, 23, 26, 28

James. Does "consciousness" exist? *Journal of Philosophy, Psychology, and Scientific Methods*, 1:477–491, 1904b. 20, 23

James. The place of affectional facts in a world of pure experience. *The Journal of Philosophy, Psychology and Scientific Methods*, 2(11):281–7, 1905. 27

James. *Pragmatism: A New Name for some Old Ways of Thinking*. Longman, Green and Company, New York, NY, 1907. 3, 4, 5, 6, 12, 18, 46

James. *A Pluralistic Universe*. Longmans, Green and Co., New York, 1909. 4

James. *Essays in Radical Empiricism*. Longman Green and Company, New York, 1912a. 46

James. The experience of activity. In James and Perry, editors, *Essays in Radical Empiricism*. Longman, Green and Company, 1912b. 4, 26

Joyce. *The Myth of Morality*. Cambridge University Press, 2001. 2

Kant. *Kritik der Reinen Vernunft*. 1781. 85

Klein. *The Rise of Empiricism: William James, Thomas Hill Green, and the Struggle Over Psychology*. PhD thesis, Indiana University, Bloomington, 2007. 24

Koskinen and Philström. Quine and pragmatism. *Transactions of the Charles S. Peirce Society*, 42(3):309–46, 2006. 47

Kripke. *Naming and Necessity*. Harvard University Press, 1980. 51, 53, 129

Lewis. *Counterfactuals*. Blackwell, Oxford, 1973. 134

Lewis. Radical interpretation. *Synthese*, 27(July-August):331–344, 1974. 9, 82

Lewis. New work for a theoryof universals. *Australasian Journal of Philosophy*, 61: 343–377, 1983. 139

Lewis. Putnam's paradox. *Australasian Journal of Philosophy*, 62:221–36, 1984. 139, 146

Lewis. *On the Plurality of Worlds*. Blackwell, Oxford, 1986. 177

Lewis. Dispositional theories of value. *Proceedings of the Aristotelian Society*, 63: 113–137, 1989. 67

Loewer. The value of truth. *Philosophical Issues*, 4:265–280, 1993. 73

Lützen. The foundation of analysis in the nineteenth century. In Hans Niels Jahnke, editor, *A History of Analysis*, volume 24 of *History of Mathematics*. American Mathematical Society, 2003. 130

Mach. *The Analysis of Sensations and the Relation of the Physical to the Psychical*. Open Court, 1984. 3, 21

Mackie. *Ethics: Inventing Right and Wrong*. Penguin, 1977. 2

Maund. The illusion theory of colour: An anti-realist theory. *Dialectica*, 60:245–68, 2006. 2

Misak. *Truth and the End of Inquiry: A Peircean Account of Truth*. Oxford University Press, Oxford, 2004. 19

Moore. *Principia Ethica*. Cambridge University Press, Cambridge, 1903. 141

Moore. Professor james's pragmatism. In Moore, editor, *Philosophical Studies*. Routledge, Kegan and Paul, 1922. 44

Parfit. *On What Matters*. Oxford University Press, Oxford, 2009. 141

Peirce. How to make our ideas clear. *Popular Science Monthly*, 12:286–302, 1878. 19

Perebroom. *Living Without Free Will*. Cambridge University Press, 2006. 2

Putnam. The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7:131–193, 1975. 127

Putnam. *The Many Faces of Realism*. The Open Court Publishing Company, 1988. 143, 145, 171

Putnam. The permanence of william james. *Bulletin of the American Academy of Arts and Sciences*, 46(3):17–31, 1992. 45

Putnam. *Ethics Without Ontology*. Harvard University Press, Cambridge, MA, 2004. 143, 171

Quine. On what there is. *Review of Metaphysics*, 2:21–38, 1948. 141, 173

Quine. Two dogmas of empiricism. In *From a Logical Point of View*. Harper Torchbooks, New York, 1953. 12, 46, 125

Quine. Variables explained away. *Proceedings of the American Philosophical Society*, 104(3):343–347, 1960. 152

Quine. *The Pursuit of Truth*. Harvard University Press, Cambridge, MA, revised edition edition, 1992. 10

Quinn. Putting rationality in its place. In *Morality and Action*. Cambridge University Press, Cambridge, 1993. 67

Rorty. Method, social science, social hope. *Canadian Journal of Philosophy*, 11: 569–588, 1981. 146

Rorty. Charles taylor on truth. In Tully, editor, *The Philosophy of Charles Taylor in Question*. Cambridge University Press, 1994a. 146

Rorty. Does academic freedom have philosophical presuppositions? In Menand, editor, *The Future of Academic Freedom*. University of Chicago Press, 1994b. 146

Royce. *The Religious Aspect of Philosophy*. Houghton Mifflin, Boston, MA, 1885. 4, 20

Russell. Pragmatism. *The Edinburgh Review*, 209:363–88, 1909. 44

Russell. The relation of sense-data to physics. *Scientia*, 16:1–27, 1914. 45

Russell. *History of Western Philosophy*. George Allen and Unwin, 1946. 45

Russell. Williams james's conception of truth. In Olin, editor, *William James: Pragmatism in Focus*. Routledge, 1992. 44

Sellars. Philosophy and the scientific image of man. In *Science, Perception and Reality*, pages 35–78. Humanities Press/Ridgeview, 1963. 42

Shope. The conditional fallacy in contemporary philosophy. *Journal of Philosophy*, 75:397–413, 1978. 132

Sider. Neo-fregeanism and quantifier variance. *Aristotelian Society Supplementary Volume 81*, 81(1):201–232, 2007. 12, 90, 93, 114

Sider. *Writing the Book of the World*. Oxford University Press, 2011. 16, 18, 137, 139, 149

Simpson. *Subsystems of Second Order Arithmetic*. Cambridge University Press, 2nd edition, 2010. 10

Sprigge. James, aboutness, and his british critics. In Ruth Anna Putnam, editor, *The Cambridge Companion to William James*. Cambridge University Press, 1997. 20

Stalnaker. A theory of conditionals. In *Studies in Logical Theory, American Philosophical Quarterly Monograph Series*, volume 2, pages 98–112. Blackwell, Oxford, 1968. 134

Stich. *The Fragmentation of Reason*. MIT Press, Cambridge, MA, 1990. 7, 49

Stich. Naturalizing epistemology: Quine, simon and the prospects for pragmatism. In Hookway and Peterson, editors, *Philosophy and Cognitive Science*, number 34 in Royal Institute of Philosophy Supplement. Royal Institute of Philosophy, 1993. 7, 60

Strong. *Why The Mind Has A Body*. Macmillan, 1903. 22

Turner. Ontological nihilism. In Bennett and Zimmerman, editors, *Oxford Studies in Metaphysics*, volume 6, pages 3–54. 2010. 151, 156, 160, 161

Walker. *The Coherence Theory of Truth: Realism, anti-realism, idealism*. Routledge, London and New York, 1989. 39

Wegner. *The Illusion of Conscious Will*. MIT Press, Cambridge, MA, 2002. 2

Weil. *The Apprenticeship of a Mathematician*. Birkhauser, 1992. 128

Weir. Neo-fregeanism: An embarrassment of riches. *Notre Dame Journal of Formal Logic*, 44(1), 2003. 116

Williams. *Descartes: The Project of Pure Inquiry*. Pelican, 1978. 136, 138

Williamson. *The Philosophy of Philosophy*. Wiley-Blackwell, Oxford, 2008. 15

Wright. *Frege's Conception of Numbers as Objects*. Aberdeen University Press, 1983. 84, 93, 123, 124

Wright. On the harmless impredicativity of $N^{=}$ (hume's principle). In Hale and Wright, editors, *The Reason's Proper Study*. Oxford University Press, 2001a. 95

Wright. On the philosophical significance of frege's theorem. In Hale and Wright, editors, *The Reason's Proper Study*. Oxford University Press, 2001b. 117

Curriculum Vitae

Thomas Mark Eden Donaldson

Education

2008 – 2012	Ph.D. in Philosophy. Rutgers, The State University of New Jersey, New Brunswick, NJ.
2007 - 2008	MMathPhil in Mathematics and Philosophy. The University of Oxford, Oxford, UK.
2004-2007	BA in Mathematics and Philosophy. The University of Oxford, Oxford, UK.

Positions

2008 – 2012	Graduate Fellow, Rutgers University.

Publications

2012	'Context Sensitivity' (with Ernest Lepore) in *The Routledge Companion to Philosophy of Language*, published by Routledge and edited by Gillian Russell and Delia Graff Fara.