

---

---

# The Diversity of Truth

A Case Study in Pluralistic Metasemantics

---

---

William Gamester

Submitted in accordance with the requirements for the degree of Doctor of Philosophy.

The University of Leeds  
School of Philosophy, Religion and History of Science

August 2017

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and William Gamester.

The right of William Gamester to be identified as Author of this work has been asserted by William Gamester in accordance with the Copyright, Designs and Patents Act 1988.

## Acknowledgements

I have a huge amount to be grateful for – I’m afraid the straight-faced sincerity necessary to properly express it doesn’t come naturally. Below is my best effort. My heartfelt thanks go to everyone on this list. It’s been the best four years, you’re all the best.

Thanks first and foremost to my supervisors. Robbie Williams has been an incredible source of wisdom and support throughout the last four years, helping to turn hunches and confusions into theories and arguments. I’m sure more or less every good idea in here has been plagiarised from an offhand comment of yours. Daniel Elstein has only come on board in the last year, but has dedicated an enormous amount of his time and energy to helping me improve the thesis. Practically every argument has been enhanced by your feedback. Paolo Santorio has seen my work as it has developed through all stages over the last four years, and has always been able to offer the most perspicuous advice for improving what I’m working on. Thanks so much to all of you for the support, patience, enthusiasm, and insight. I’m still quite overawed to have had such brilliant philosophers to go over my ideas and arguments, and to take them seriously. Granted, you were contractually obliged to do so, but it’s still been a wonderful experience for which I am grateful beyond words.

I’d also like to thank the rest of the staff in the Leeds philosophy department for their contributions towards my philosophical development, be it in questions after talks or conversations in pubs. Pekka Väyrynen, in particular, has given up his time on more than one occasion, most notably when I was first finding my feet in the thick ethical concepts literature. I’d also like to thank my examiners, Matti Eklund and Jack Woods, for an invigorating *viva voce*.

I’ve been lucky enough to have been a part of three fantastic philosophical groups over the course of my Ph.D. Perhaps most important is the Leeds postgraduate philosophy community. It’s been a blast. Thanks to Daniel Abrahams, Dani Adams, Sarah Adams, John Baldari, Mike Bench-Capon, Thomas Brouwer, Sam Carelmon, Fabio Ceravolo, Pei-Lung Cheng, Marc Cole, Adina Covaci, John Fernley, Jade Fletcher, David Heering, Lewis Hickley, Olof Leffler, Gary Mullen, Alice Murphy, Simon Newey, Christina Nick, Tadgh Ó Laoghaire, John Parry, Adam Patel-Summers, Emily Paul, Mat Rees (of the famed Rees-Gamester Hypothesis), James Rimmer, Bryan Ross, Sam Symons, Nahuel Sznajderhaus, Nick Tasker, Alison Toop, Andrea White, Marc Wilcox, and Nathan Wood – I don’t say this lightly, but you’re all basically tolerable as human beings. Thanks for making the last four years pretty much ok. In places. You’re alright.

Special thanks are due to Jade Fletcher and Mike Bench-Capon for innumerable ultra-helpful chats about truth over the years, and to Nick Tasker for four years of weird philosophical brotherhood. You guys, in particular, are alright.

Second is the international pluralisms community. I’ve had great trips to the University of Connecticut, U.S.A., Yonsei University, South Korea, and the University of Bologna, Italy, all to talk about this niche area of philosophy of language and logic (including a workshop on my thesis in Bologna). It’s really been fantastic. Thanks to Douglas Edwards, Filippo Ferrari, Nathan Kellen,

Jiwon Kim, Junyeol Kim, Teresa Kouri Kissel, Sebastiano Moruzzi, Nikolaj Pedersen, Paul Simard Smith, Erik Stei, Elena Tassoni, Pilar Terrés Villalonga, Cory Wright, Jeremy Wyatt, Andy Yu, and Luca Zanetti for the good times. And the philosophy. I don't know about this "pluralism" malarkey, but travelling around the world and hanging out with you all has been a lot of fun. You're alright.

Third is the Nature of Representation research group at the University of Leeds. Not many people get this kind of opportunity – I've learned so much, not just from the weekly seminar, but the near-constant conferences on awesome topics and visiting speakers. As well as Robbie, Daniel, and Nick, my thanks go to the postdocs, Jennifer Carr, Ed Elliott, Rachel Goodman, Simon Hewitt, and Gail Leckie. I'm yet to hear if we solved the nature of representation, I assume we did – if we didn't manage it between us I doubt it'll ever get done. It's been great working with you all. You're alright. And thanks to the ERC for funding the project and my research. You're alright.

Many of the chapters here have been presented in talks and seminars, and many people have provided useful feedback on these occasions or on drafts of papers. In addition to those mentioned above, I'd like to thank: Giulia Casini, Paul Horwich, Jess Isserow, David Liggins, Michael Lynch, and James Woodbridge. You're alright.

I have been extraordinarily lucky throughout the tumultuous highs and lows of postgraduate life to have had a veritable mob of friends and family around me. Frankly, the almost embarrassing levels of love and support has at times bordered on the irritating. Every once in a while, you want the whole world to abandon you. But, no. Thanks to all, and particularly to the following.

Thanks to my housemates. Chronologically: Allie Rinck, Anna Heaton, Joey Whitfield, Ellis Jones, Bex Jarman, Daniel Ceballos Contreras, and (at the end now) Simon Lowe. Living with you all at the Grove has been the best – you've really made home feel like a home away from home. Thanks. You're alright.

Mum and Dad. You have rather selfishly put me in the awkward position of having to try and express gratitude for over 28 years of unconditional love and support in just a few sentences. If I don't believe in me, your belief in me makes me believe in me, even though I know that I have better evidence about me than you do. Thanks for everything, for always being there. You're alright.

Thanks are also due to Peter Gibson and especially my brother, Ed, for inspiring me to take up philosophy. In fact, pretty much everything I'm interested in I've copied off Ed. Thanks for always putting up with that, and being the strongest (mentally and physically) and best person I know, a kaleidoscopic commotion of neon colour in a grey world. Thanks for the lifetime's friendship and making me a better person. You're alright.

And thanks finally to my best friend and partner, Emma. You've been there with sympathy and patience as I've oscillated between despair and exhilaration over esoteric things you couldn't care less about. Thanks for being there through it all. I'll try not to forget you when the mainstream success of this thesis inevitably leads to an international movie franchise. I don't know how I'll cope without you. Badly, I expect. Thanks for being the best bit of life, every day. You're alright.

## Abstract

This thesis concerns pluralism about truth: roughly, the theory that there is more than one way to be true. Where ‘Grass is green’ might be true in one way, ‘Eating meat is wrong’ or ‘ $7 > 3$ ’ might be true in another. I am interested in showing this theory in its best light. This requires casting a critical eye over extant incarnations of pluralism, formulating new, stronger motivations in its favour, and defending it from objections.

Where most pluralists try to motivate the theory by assuming an underlying *ontological* diversity – in what different truthbearers are about, e.g., grass vs. wrongness vs. numbers – my arguments assume an underlying diversity, not in the world, but in our thought and talk. While ordinary discourse like ‘Grass is green’ expresses representational states (the belief that grass is green), I assume with metaethical expressivism that moral discourse like ‘Eating meat is wrong’ expresses desire-like states (e.g., disapproval of eating meat). Given this metasemantic pluralism, I provide a direct argument for thinking that truth within ordinary discourse consists in corresponding with reality, while moral truth is epistemically constrained; and I develop a novel theory of moral truth.

I go on to argue that the most prominent objections to pluralism – which concern cases where truthbearers apt for different properties are “mixed” together – in fact pose no special problems for the pluralist. I provide a pluralist-friendly metaphysics of truth for complex truthbearers that dissolves the appearance of difficulty, arguing that the truth of a complex consists in a distinct property that is grounded in the truth properties relevant for its components. And in the final chapter, I show how this independently motivated metaphysics of truth can in turn be used to dissolve the liar paradox.

# Table of Contents

Title Page	<i>i</i>
Acknowledgements	<i>iii</i>
Abstract	<i>v</i>
Table of Contents	<i>vi</i>
Table of Figures	<i>x</i>
<b>Introduction</b>	<b>1</b>
<b>Part I – Pluralism about Truth</b>	<b>4</b>
<b>Chapter 1: Truth and Pluralism</b>	<b>5</b>
1.1    Language, Thought, Reality	5
1.1.1    Predicate, Concept, Property	5
1.1.2    Truthbearers	9
1.2    Deflationism and Inflationism	11
1.2.1    Deflationism	11
1.2.2    Realist Inflationism	15
1.2.3    Anti-Realist Inflationism	17
1.3    Pluralisms	19
1.3.1    <‘Truth’>: Pluralism, Monism, Inconsistency	19
1.3.2    A Panoply of Property Pluralisms	22
1.4    Methodology: Analysis and Engineering	28
1.4.1    Role-First Methodology	28
1.4.2    From Concept Monism to Property Monism?	31
Conclusion	33
<b>Chapter 2: Scoping Out Truth Pluralism</b>	<b>34</b>
2.1    Proto-Scope Problems	34
2.1.1    Proto-Scope Problems	34
2.1.2    Explaining Anti-/Realist Intuitions	37
2.2    Ontologically-Driven Scope Problems	40
2.2.1    Cotnoir & Edwards: Truth, Being, & Grounding	40
2.2.2    Pedersen: Mind In/dependence and Metaphysical In/completeness	43

2.2.3	Lynch: Causal im/potence and Mind In/dependence	46
2.3	A New Strategy	50
<b>Part II – A Metasemantically-Driven Argument for Truth Pluralism</b>		<b>52</b>
<b>Chapter 3: Explanation, Success, and Coincidence</b>		<b>53</b>
3.1	Deflating Success	54
3.1.1	The Success Argument	54
3.1.2	Individualised Explanations	57
3.1.3	Truth’s Expressive Role	59
3.2	A Problem Case	60
3.2.1	Coincidental Success	60
3.2.2	The Threat of Collapse	62
3.3	Twisting the Knife	65
3.3.1	Supplementing the Explanation?: Explaining the Premises	65
3.3.2	Supplementing the Explanation?: Further Premises	69
3.3.3	The Inadequacy of Mere Generalisation	71
	Conclusion	74
<b>Chapter 4: Expressivism and Truth</b>		<b>76</b>
4.1	The Scope of Success	76
4.1.1	Correspondence and Success	76
4.1.2	The Scope of the Success Argument	79
4.2	Metaethical Expressivism, Moral Truth, and the Frege-Geach Problem	79
4.2.1	Metaethical Expressivism	79
4.2.2	The Frege-Geach Problem	81
4.2.3	‘A Panoply of Dignities’	84
4.2.4	A “Fast-Track” Solution?	88
4.3	Expressivism and the Normativity of Truth	92
4.3.1	A Normative Role for Truth	92
4.3.2	The Normativity of Truth for Belief	94
	Conclusion	96
<b>Chapter 5: Why Expressivists Can Only Go So Far Wrong</b>		<b>97</b>
5.1	Fallibility and (Super-)Stability	97
5.1.1	Fallibility for Expressivists	97
5.1.2	Egan’s Challenge	100

5.1.3	Improving Improvement	101
5.1.4	Super-Stability and Inaccessibility	104
5.2	Tightening the Net	105
5.2.1	Anti-Scepticism	105
5.2.2	Weak Super-Stability and Inaccessibility	107
5.2.3	Divergent Improvements	108
5.2.4	True Divergence as Scepticism	111
5.2.5	The Necessity and Sufficiency of Weak Super-Stability for Truth	112
5.3	“Smugness”	114
5.3.1	The Superficial Smugness Worry	114
5.3.2	Establishing a Network – Some Formal Results	115
5.3.3	Against True Divergence	116
5.4	Moral Truth as Weak Super-Stability	118
	Conclusion	120

### **PART III – LOGIC, LIARS, AND LOGICAL FORM** 121

#### **Chapter 6: Logic, Logical Form, and the Disunity of Truth** 122

6.1	“Mixing” Problems?	123
6.1.1	The Problem of “Mixed” Inferences	123
6.1.2	The Problem of “Mixed” Compounds	127
6.1.3	Against Moderate Pluralism	132
6.2	Form-Restricted Pluralism	135
6.2.1	The Theory	135
6.2.2	Objections and Replies	138
	Conclusion	141

#### **Chapter 7: Disentangling Truths** 143

7.1	The Problem of Mixed Atomics	143
7.1.1	Domain Individuation	143
7.1.2	Mixed Predicates	146
7.2	Disentangling Truths	148
7.2.1	Expressivist Approaches to Thick Predicates	148
7.2.2	Disentangling Truths	151
	Conclusion	154

#### **Chapter 8: Truth as None and Many** 155



8.1	Generalisation and the Liar	155
8.1.1	The Problem of Mixed Generalisations	155
8.1.2	The Liar and Inconsistency	157
8.2	Why Infinite Liars are Better than One	159
8.2.1	Tarski and Cotnoir on Infinite Liars	159
8.2.2	The Infinite Liars of Form-Restricted Pluralism	161
8.2.2.1	A First Case	161
8.2.2.2	A Problem Case	162
8.2.2.3	Generalisation the Solution	166
8.2.3	Revenge?	167
	Conclusion	168
	<b>Concluding Remarks</b>	170
	References	173

## Table of Figures

Figure 1	Flowchart of Truth Pluralisms	26
Figure 2	(Divergent) Improvements	109

---

# INTRODUCTION

---

The job of semantics is to tell us what the semantic properties (referents, extensions, intensions, characters, truth conditions, etc.) of linguistic entities are. The job of metasemantics is to explain why linguistic entities have the semantic properties that they have. A semantic theory will tell you that the word ‘Ajax’ refers to my cat, Ajax. A metasemantic theory will tell you *why*; why it is that *that* word refers to *that* cat. By a “pluralistic metasemantics”, I mean a metasemantic theory that tells a different story for different regions of discourse.

Other things being equal, explanatory unity is a theoretical virtue. But some phenomena are complex enough that they do not permit of unified explanations. Pluralistic metasemantics has it that language is one such phenomenon. The later Wittgenstein once compared the study of language to

‘...looking into the cabin of a locomotive. There are handles there, all looking more or less alike. (This stands to reason, since they are all supposed to be handled.) But one is the handle of a crank, which can be moved continuously (it regulates the opening of a valve); another is the handle of a switch, which has only two operative positions: it is either off or on; a third is the handle of a brake-lever, the harder one pulls on it, the harder the braking; a fourth, the handle of a pump: it has an effect only so long as it is moved to and fro.’ (2009: §12).

At the heart of Wittgenstein’s simile is not just the idea that the “handles” of language may all function differently, but that they might do so despite “all looking more or less alike” – that is, despite a similarity in surface appearances. (This similarity itself being no mere coincidence, but functionally useful.)

Consider the following two sentences:

- (1) Eating meat is normal.
- (2) Eating meat is wrong.

On the face of it, (1) and (2) are very similar. Their surface grammar is identical: each is formed by concatenating one and the same singular term, ‘eating meat’, with a monadic predicate, ‘is normal’ and ‘is wrong’, respectively. Both can be embedded meaningfully into various other sentences, into *inter alia* logical compounds, belief-attributions, and truth-attributions. Both can appear in valid inferences. And both are (let’s suppose) true. The vast majority of people in the U.K

are carnivores, and yet farm animals are often kept in abhorrent conditions, and the livestock industry contributes more to greenhouse gas emissions than all (other) transport combined.

But for all their surface similarities, there is a crucial difference. Where (1) only describes how things *are*, (2) says how things *ought to be*. Where (1) merely ascribes some property to eating meat, (2) hereby condemns it as something that *ought not* to be done. In short, (2) is normative where (1) is merely descriptive.

Metaethical expressivism,<sup>1</sup> as I understand it, claims that the semantic properties of a moral sentence like (2) are to be explained in terms of some conative or affective, or generally desire-like, mental state that it *expresses*; e.g., (2) might express *disapproval* of eating meat. Ordinary descriptive sentences like (1), by contrast, express representational states; i.e., the *belief* that eating meat is normal. If this is right, then the metasemantics of moral discourse is different from the metasemantics of ordinary descriptive discourse. (I will go through this contrast in much more detail in Chapter 4.)

My concern in this thesis is with the nature of *truth* as a property of intentional entities – sentences, propositions, or beliefs. Part II (Chapters 3-5) provides a detailed argument to the effect that, given this kind of metasemantic pluralism, we ought to think that the nature of truth *varies* between those regions of discourse that permit of different metasemantic explanations. In Chapters 3 and 4, I argue that truth must consist in a substantive relation that holds between the relevant truthbearer and that part of reality that it is about – or *correspondence* – if we are to explain the practical success of actions performed by rational agents.<sup>2</sup> But in Chapter 4, I set out metaethical expressivism in some detail, and argue that moral truths, expressivistically-conceived, fall outside the scope of this argument. I then argue in Chapter 5 that expressivists are committed to thinking that the moral truths are epistemically constrained, and I develop a new anti-realist theory of truth – which I call *weak super-stability* – that explains this constraint. While the truth of (1) might consist in correspondence, the truth of (2) consists in weak super-stability.

This kind of pluralism about the nature of truth is supposed to face a number of embarrassing problems the moment truthbearers from different discourses are “mixed” together, in inferences, logical compounds, and generalisations. In Part III (Chapters 6-8), I aim to show that this impression is mistaken: far from creating new problems, pluralism about truth provides new solutions to old problems. In Chapter 6, I show that the so-called Problems of “Mixed” Inferences and Compounds are really just instances of perfectly general, structural constraints on any metaphysics of truth, which monism *qua* monism doesn’t automatically satisfy. I then detail and defend a (pluralist-friendly) metaphysics of truth for logically complex truthbearers that satisfies these constraints. The theory postulates a further variety of pluralism about truth: that the truth of a complex truthbearer consists in a property distinct from, but grounded in, the truth properties

---

<sup>1</sup> Or, more generally, metanormative expressivism. I restrict my attention to the metaethical case throughout the thesis for simplicity.

<sup>2</sup> Strictly, this argument is only shown to apply to *some* representationalist discourse; see §4.1.2.

relevant for its components. In Chapter 7, I offer a new and principled response to the Problem of Mixed Atomics using this theory.

Finally, in Chapter 8, I show how the metaphysics of truth defended in Chapters 6 and 7 can – in conjunction with an inconsistency theory of the *concept* of truth – be used to dissolve the liar paradox, by developing a proposal structurally analogous to the orthodox Tarskian hierarchical solution, but using only machinery that is independently motivated within the thesis. The aim is to hereby resolve the tension between the arguments of Part II, which aim to show that truth is an explanatorily and normatively important part of our world, with how readily truth seems to give rise to paradox.

The resultant view is complex, and while many of the primary elements have been defended before, they have never (to my knowledge, anyway) appeared in this combination. To make my case as strong as possible, I intend to start from a comparatively neutral starting point, trying to take on as few controversial assumptions as possible besides the core metasemantic pluralism (and this only from Chapter 4 onwards). Doing so requires some stage-setting, which comes in Part I (Chapters 1 and 2). Chapter 1 introduces most of the various elements that will make up the whole. It also concerns questions of methodology (I take mine to be an instance of what is contemporaneously called “conceptual engineering”) and asks what it might mean to be a “pluralist” about truth, formulating a complex taxonomy of possible views, many of which seem to be systematically overlooked in the literature. Chapter 2 looks at extant motivations for the kind of “discourse-dependent” truth pluralism I’m interested in. These arguments are typically “ontologically-driven” – they appeal to differences in the nature of those things different discourses are *about*. I find them for the most part wanting. This sets the stage for my own “metasemantically-driven” argument in Part II.

---

---

PART I

PLURALISM ABOUT TRUTH

---

---

# CHAPTER 1

## TRUTH AND PLURALISM

---

‘I am inclined to think that pluralism should be the default view for most philosophical expressions. Typically there will be no single privileged role associated with such an expression, and different roles will be played best by different properties.’

Chalmers (2011: 540).

The primary purpose of this chapter is to set the stage for the rest of the thesis by introducing most of the various elements that will ultimately make up the theory as a whole, and to thereby situate my position with respect to the literature. Section 1 distinguishes between the predicate, concept, and property of truth, and addresses the related question of what kinds of thing can be true in the sense we are interested in here; i.e., what *truthbearers* are. Section 2 sets out deflationary and both realist and anti-realist inflationary views on the nature of truth. In Section 3, I aim to provide the most sophisticated taxonomy of pluralist views about truth to date. This is no mere exercise in logical space mongering: it strikes me that various argumentative moves made in the literature trade on overlooking viable positions distinguished here.

It is natural to think that investigation into the concept of truth must precede investigation into the property, since the concept identifies the property of interest. In Section 4, I explain why I reject this “concept-first” methodology in favour of a “role-first” methodology, which I take to be an instance of what is contemporaneously called “conceptual engineering”.

### 1.1 Language, Thought, Reality

#### 1.1.1 *Predicate, Concept, Property*

Predicates I take to be linguistic entities – specifically, sentence-forming operators on singular terms – which we will in general denote using single-quotes.<sup>1</sup> The predicate ‘is a cat’, for example, is an expression of the English language. (I thus work with an intuitive notion of word individuation, such that words are not individuated by their meanings; the French ‘est un chat’ is a different predicate to the English ‘is a cat’.) The *extension* of a predicate is the set of things that it

---

<sup>1</sup> Throughout the thesis, I will use single-quotes to denote linguistic entities and to quote authors, while double-quotes are reserved for scare-quoting and other “half-use-half-mention” functions.

applies to, that it can be truly predicated of. In the actual world, the extension of ‘is a cat’ is the set of all and only the cats that actually exist. (One might alternatively think of extensions as functions from things to truth-values; the extension of ‘is a cat’ is then a function which returns ‘true’ if the thing is a cat, and ‘false’ otherwise.) The *intension* of a predicate is a function from (perhaps centred) possible worlds and contexts of evaluation to extensions. At each possible world, the intension of ‘is a cat’ returns the set of things that are cats in that world. Context-sensitive words have *characters*, which are functions from contexts of utterance to intensions. ‘Here’, for example, returns the location of the utterance. The *referent* of a monadic predicate is a property, which we will say it *ascribes*.<sup>2</sup> (Polyadic predicates ascribe relations; the adicity of a predicate is determined by the number of singular terms it operates on to form a sentence.) A predicate’s corresponding singular term (for ‘is a cat’, something like ‘cathood’ or ‘catness’ (not to be confused with the Hunger Games character); for ‘is true’, ‘truth’) *denotes* the property the predicate ascribes. Exactly what a word is is a matter of ongoing dispute, about which I intend to be neutral beyond what I have said here.

A concept <F>, which we will denote with angle-brackets,<sup>3</sup> is first and foremost (though somewhat roughly speaking) what someone possesses when she is able to think of something as being F. So, someone possesses the concept <cat> just in case she is able to think of something as being a cat. We will also say that the concept <F> is what someone “grasps” when she understands and is a competent user of the word ‘F’; someone who understands ‘is a cat’ grasps the concept <cat>. We will say that a word *voices* its associated concept.<sup>4</sup> A concept <F> is distinct from a concept <G> just in case it’s possible for someone to rationally think of something as being F without thinking of it as being G, or vice versa. There is also an ongoing dispute about the nature of concepts: they may be mental particulars, abilities, or abstract entities akin to Fregean senses.<sup>5</sup> We won’t take a stand on this, but as a working model I typically think of concepts as kinds of mental particular, something like the mental analogue of words (though without being committed to anything as strong as the Language of Thought Hypothesis).

Someone who grasps a concept will have an associated *conception*, which is roughly a set of beliefs associated with the concept. For instance, someone living 1,000 years ago who grasped the concept <water> would have in her associated conception that water is a clear, potable liquid, which slakes thirst, and which boils when sufficient heat is applied to it. But it would not include, as our conceptions include, that at the molecular level water contains hydrogen and oxygen. The rough idea here is that concepts are public and shareable where conceptions can be idiosyncratic. Indeed, given familiar externalist arguments (Putnam 1975, Burge 1979), we can assume that, in at least some

---

<sup>2</sup> Context-sensitive predicates ascribe different properties in different contexts.

<sup>3</sup> I have nothing against the small-caps convention other than that it is time-consuming.

<sup>4</sup> It is more usual to say that words *express* concepts, but following Gibbard (2012: 3) I want to reserve the word ‘express’ for the relation between sentences and mental states outlined below.

<sup>5</sup> See the papers collected in Margolis & Laurence (1999) for an overview, especially Laurence & Margolis (1999).



cases, someone can possess a concept despite having a (quite radically) *mistaken* associated conception.

Properties, as mentioned, are the “worldly” referents of predicates, which we will denote using the relevant singular terms. I thus work with an abundant conception of properties (and relations) in Lewis’s (1983) sense, although with Lewis I assume that a proper subset of properties abundantly construed ground objective similarity relations and are explanatorily potent, and are thus “sparse”.<sup>6</sup> Objects *exemplify* properties; where the predicate and concept can be used to say or think of something that it is a cat, Ajax the cat himself exemplifies the property of being a cat, regardless of what anyone says or thinks about him. (Though there may well be mind-dependent properties, such that an object’s exemplification of that property does in some sense depend on what people say or think.) Properties thus determine intensions, and are therefore more finely individuated than their actual extensions.<sup>7</sup> Even if every vertebrate with a heart is a vertebrate with a liver,<sup>8</sup> the property of *being a vertebrate with a heart* is distinct from the property of *being a vertebrate with a liver* because it’s possible that there should be a creature that exemplifies one without exemplifying the other. I will assume that necessarily co-extensional properties are identical. Exactly what properties are is also a matter of ongoing dispute, about which I intend to be neutral beyond what I have said here.

On the above test for concept individuation, concepts are more finely individuated than properties. To use the ubiquitous example, one can rationally think that the liquid in the glass is water without thinking that it is H<sub>2</sub>O if, for instance, one does not know that water is H<sub>2</sub>O. Concepts thus play the role of Fregean senses here (though we need not conceive of them as Frege did). The concept a word voices determines at least in part<sup>9</sup> its semantic features; its extension, intension, and referent. We might say that the concept voiced by a context-sensitive word determines in part its character; but it seems preferable to say that the concept voiced is determined in part by its character, so you and I voice different concepts with the word ‘me’ (or the choice may vary between context-sensitive words). This won’t especially matter for our purposes.

Distinguishing the concept of truth from the property of truth allows us to individuate two broad and distinct subject matters. Inquiry into the nature of the property of truth involves trying to say, in an informative way, what the nature of that property consists in; what property it is that something possesses when it is true. Inquiry into the concept <truth>, by contrast, might involve

---

<sup>6</sup> If one thinks that only sparse properties are *really* properties, then one is welcome to translate the thesis into terms one prefers. Nothing turns on the terminological choice.

<sup>7</sup> Strictly, for *properties* to be co-extensional is really for the predicates that ascribe them to be co-extensional. I will sometimes be sloppy on this front; I will also talk of concepts ascribing and denoting properties, but this should be read as shorthand for the predicates which voice them doing so.

<sup>8</sup> Which is so according to Kirkham (1992: 4-5).

<sup>9</sup> Perhaps, e.g., in conjunction with features of the external environment; though one may favour “wide” individuation of concepts, such that it includes such features. So, one may want to say that Oscar and Twin Oscar voice the same concept with ‘water’, though they denote different properties; or one may include the property denoted in the individuation of the concept. I lean towards the former, “narrow”, individuation.

trying to say, in an informative way, what competence with the word ‘is true’ consists in; what it is that someone grasps when they grasp concept <truth>. <sup>10</sup>

Let me emphasise from the outset that conceptual simplicity or complexity need not track metaphysical simplicity or complexity. <Water> is presumably a comparatively simple concept to master (simpler, in any case, than <H<sub>2</sub>O>), and yet *water* has a complex nature, which it took thousands of years of scientific progress to uncover. Now, it is highly plausible that some concepts have complex structures. For instance, <vixen> might be a complex concept composed out of <female> and <fox>. (<H<sub>2</sub>O> out of <hydrogen> and <oxygen>.) It is very plausible that there are *some* such concepts because we can “build” them. To use an example that will be important later (§7.1):

*Badwork*

$\forall x(x \text{ is badwork} \leftrightarrow (x \text{ is time-consuming} \ \& \ x \text{ is wrong}))$ .

<Badwork> is a complex concept composed out of <time-consuming> and <wrong>. (I don’t pretend the nature of this “composition” is obvious; it will depend, ultimately, on one’s account of concepts. But it is quite common to think of concepts as having structure of some kind. See, e.g., Laurence & Margolis 1999.) For such concepts, we may be able to offer a “decomposing” analysis, which deconstructs the concept to its constituents. And in some cases this will track metaphysical complexity: *being a vixen* is arguably grounded in *being female* and *being a fox*. But such a process must have an end, so there must be some terms such that our competence with them needs to be explained some other way;<sup>11</sup> and these will *not* be those concepts that denote the metaphysically “simple” properties, if such there be (whatever exactly metaphysical simplicity amounts to). The top contenders here are highly theoretical properties, like charge and spin, which take great conceptual sophistication to understand.

Those who offer sophisticated theories as analyses of the *concept* of truth face the following worry:

“The vast majority of the work on the nature of truth is engaged in trying to find a conceptual analysis of truth. This would be something like a definition of truth in terms that are more primitive or more fundamental or better understood or less controversial. Many have

---

<sup>10</sup> The former is thus akin to what Haslanger (1999, 2000, 2005, 2006) calls a “descriptive” or “externalist” project, focusing on what she calls the “operative concept”; the latter to her “conceptual” or “internalist” project, focusing on the “manifest concept”.

<sup>11</sup> Compare the metaethical expressivist, who thinks that ‘good’ voices a simple concept, but allows for substantive first-order normative theories as to the nature of goodness. The possibility of an alethic expressivism is a fascinating one, which I think can be made consistent with large swathes of what is argued in this thesis, but which I unfortunately do not have the space to develop here. See, e.g., Schroeder (2010a), Gibbard (2012: ch.7), Ridge (2014: ch.7).

criticized this kind of project before, mostly on the grounds that there is nothing more primitive, fundamental, better understood, or less controversial.’ (Scharp 2013: 266)

But it is quite compatible with <truth> being a comparatively primitive or fundamental concept that it denotes a complex, interesting property. Such theories are thus often better construed as being at the level of property than concept, and I will charitably interpret them as being so.<sup>12</sup>

### 1.1.2 Truthbearers

Both ‘true’ and ‘truth’ (or ‘being true’) are polysemous in our language. You can be true to something, like a person or a cause. Someone can be your one true friend, or your true love. An arrow can fly true. ‘Truth’ can sometimes be a mass noun (‘There was a lot of truth in what she said’) or a count noun (‘Six truths were written on the board’). We are interested in truth as a property of intentional entities, things which are *about* things; things like beliefs, propositions, sentences, and utterances.<sup>13</sup> Call these *truthbearers*.

Sentences I also take to be linguistic entities, composed out of words, and so we will also denote them with single-quotes. ‘Ajax is a cat’ is the sentence composed out of the singular term ‘Ajax’ and the predicate ‘is a cat’. Utterances are uses (as opposed to mentions) of sentences. Beliefs are mental states that agents can be in. (We’ll say more about the nature of belief in Chapter 4.) Sentences can be used to *express* mental states; and, in particular, declarative sentences can be used to express beliefs. A sincere utterance of ‘Ajax is a cat’, for instance, expresses the belief that Ajax is a cat.

Things are rather more complicated with propositions, which (as thought-associated things) we will also denote with angle-brackets. Propositions, the lore has it, are the objects of belief and other so-called “propositional attitudes”, like desire, hope, supposition, wonder, fear, and so on. That is, to a first approximation: belief is a two-place relation which a believer stands in to a

---

<sup>12</sup> Partly for these reasons, I confess that I do not understand “primitivism”, the view associated with Moore, Frege, and more recently Davidson (1990, 1996), and defended of late by Patterson (2010) and Asay (2013). It typically sounds like the view that the *concept* does not “decompose” into other, more basic, concepts; but that, I think, that can be accepted by most participants in the debate. It sometimes sounds like the view that our semantic competence with ‘is true’ cannot be explained *at all*, but this is a bizarre view. (Even a “natalist” who thinks <truth> is innate would presumably think there’s an explanation of our capacity to think with it.) I do not see what, other than despair, could licence refusing to offer any explanation whatsoever. (And as a view about the *property*, it would either amount to an implausible non-naturalism or fundamentalism (mentioned later), or again sheer pessimism.) Patterson’s primitivist, for example, thinks that <truth> is circular, and he responds to the worry that this ‘leaves its intension unsettled to an unacceptable degree.’ (2010: 22) But *that* is not what is wrong with circularity in this context; it’s that it’s utterly uninformative. Patterson’s primitivist explains nothing, and what’s the point in theories that don’t explain anything? (Asay’s view, which tries to combine primitivism about <truth> with (i) deflationism about the property, (ii) the explanatory potency of truth, and (iii) an inflationary theory of truthmakers, is a minefield.)

<sup>13</sup> This is on the face of it a case of true polysemy, rather than mere ambiguity, as one finds with, e.g., ‘coach’, which is ambiguous between *bus* and *instructor*; the meanings of the words seem systematically linked. Daniel Elstein suggests to me that the common thread is *faithfulness*.

proposition, which either is or determines the *content* of that belief. <Ajax is a cat> thus either is or determines the content of the belief that Ajax is a cat.

However, propositions are *also* supposed to be (or determine) the “meanings” of sentences; we will say that sentences *voice* propositions.<sup>14</sup> Now, the extensions of the words in a sentence together determine its truth value; their intensions determine a set of possible worlds at which that sentence is true (so it is true just in case the actual world is in that set). Call this set the sentence’s *truth conditions*. The structured entity composed out of the referents of the words in the sentence we will call its *Russellian proposition*. The structured entity composed out of the concepts voiced by the words we will call its *Fregean proposition*.<sup>15</sup> When we say that propositions are the meanings of sentences, we might mean that they are sets of possible worlds (truth conditions), Russellian propositions, or Fregean propositions. Or, perhaps, something else again. Scott Soames, for instance, takes propositions to be types of cognitive event: the proposition <Ajax is a cat> is the type of event that is tokened when I mentally predicate catness of Ajax.<sup>16</sup> (It is also quite possible that the two theoretical roles mentioned above – the objects of belief and the meanings of sentences – come apart.)

On the face of it, then, we have a number of different topics of interest: doxastic-truth, propositional-truth, sentential-truth, utterance-truth.<sup>17</sup> This gives us our first monist/pluralist division, though one of comparatively little interest. At the level of property, the monist would hold that every different kind of truthbearer, when true, possesses the same property. But this is unattractive, insofar as the ontological diversity of beliefs, propositions, and sentences makes identification of such a property problematic.<sup>18</sup> And, in any case, the systematic connection between them makes this option unappealing. It is more usual to take a pluralistic approach here, and offer different accounts of truth for different kinds of truthbearer. One might, in theory, offer an account for a particular kind of truthbearer while remaining neutral on the others, but the usual pluralistic approach is to take one kind of truthbearer to be *primary*, such that the nature of truth for the other truthbearers is to be understood in terms of the truth of the privileged kind. Call this approach *primalism*.

The default such view says that propositions are the primary truthbearer. (Indeed this is typically a stipulated theoretical role of propositions.) A belief is true just in case its content is a true proposition; a sentence is true just in case it voices a true proposition; an utterance is true just in case

---

<sup>14</sup> Again, ‘express’ is more usual here; see fn.4.

<sup>15</sup> We need not think that *each* word in a sentence contributes its actual referent or concept to the proposition voiced. If ‘long’ appears as part of expression ‘long hard stare’, then it may determine a referent or concept in conjunction with ‘hard’ and ‘stare’, and the output may be the component of the proposition. Propositions may also contain components that are not contributed by any particular expression in the sentence, but instead by context or whatever.

<sup>16</sup> See Soames 2010; King, Soames, & Speaks 2014.

<sup>17</sup> For ease of presentation, I am ignoring, e.g., assessment-sensitive views (e.g., MacFarlane 2014), on which truth is a dyadic property of propositions *at a context of assessment*.

<sup>18</sup> This might seem far from problematic for the deflationist. Deflationists also typically opt for primary truthbearers, however.

it is the use of a true sentence. Giving an account of the nature of truth for propositions thus derivatively gives us our account of truth for other truthbearers. Analogous views can be formulated whichever truthbearer is taken to be primary. For illustration, take the following version of the Language of Thought hypothesis. The belief that Ajax is a cat, say, involves having the sentence ‘Ajax is a cat’ – but “written” in the language of thought (“Mentalese”) – tokened in your brain, such that it plays the belief-role. One might then take Mentalese sentences to be the primary truthbearer: a belief is true just in case the relevant Mentalese sentence is true; a public language sentence is true just in case it expresses a true belief. This does without propositions altogether.

Note that primalism is often formulated at the level of *concept*. This involves attributing a further polysemy to ‘true’ even within its application to intentional entities: that it voices one concept  $\langle \text{truth}_{\text{prop}} \rangle$  when applied to propositions, but another  $\langle \text{truth}_{\text{sent}} \rangle$  when applied to sentences; and that  $\langle \text{truth}_{\text{sent}} \rangle$  is to be analysed as:  $\langle \text{voices a true}_{\text{prop}} \text{ proposition} \rangle$ .

While striving to be as neutral as possible, by default I will work with the property primalism outlined above, assuming that propositions are the primary truthbearer. In the end, I will only be committed to two claims with regards to truthbearers: (i) that some sentences are non-primary; and (ii) that primary truthbearers have (sentential) structure. (See especially §7.2.) So, if propositions *are* the primary truthbearers, then they are not just sets of possible worlds. (Note that this is compatible with the meanings of sentences being just truth conditions; the meaning of a sentence is then just not the proposition that it voices, though it is determined by the proposition it voices. Similarly if the “content” of a belief is just its truth conditions: then belief is a relation to truth conditions *via* a proposition, which determines those truth conditions.) In discussion, I will either talk neutrally in terms of “truthbearers”, or “propositions” while remaining neutral on what these are; I will then offer clarifications in the footnotes when the nature of the truthbearers is important.

## 1.2 Deflationism and Inflationism

### 1.2.1 Deflationism

In Part II, I argue against deflationism in favour of a realist truth property for some truthbearers, and an anti-realist property for others. In this section, I introduce deflationism on the one hand, and realist and anti-realist truth properties on the other.

In my view, truth is overburdened: there is nothing in the world that does all the work that we expect truth to do. This may sound surprising, given that over the last 70 years many philosophers of truth have spent a lot of time arguing that truth, qua property, is in fact not very interesting *at all*. On the contrary, so-called *deflationists* about truth maintain that there is nothing interesting to say about what truth *is* – about the *nature* of truth, or what truth *consists in* – but instead argue that the only interesting question concerns what our mastery of the concept of truth, or our capacity to use

the truth predicate, consists in; a question to which deflationists have offered a dizzying array of answers. Here, the deflationist camp divides into *inter alia* disquotationalists, minimalists, inferentialists, modest theorists, and prosententialists. It will be useful to briefly outline these most prominent varieties.

Disquotationalists like the later Hartry Field emphasise the disquotational schema (DisS), while minimalists like Paul Horwich emphasise the denominalisation schema (DenS), the primary difference being whether they take sentences or propositions to be the primary truthbearers:

$$\begin{aligned} (\text{DisS}) \quad & \text{'p' is true iff } p. \\ (\text{DenS}) \quad & \langle p \rangle \text{ is true iff } p.^{19} \end{aligned}$$

Where for 'p' we can substitute a declarative sentence. For Field (1994a, 1994b), the core claim is that the right-hand side of any instance of (DisS) is “cognitively equivalent” to the left-hand side, in the sense that a ‘person’s inferential rules license (or, license fairly directly) the inference from one to the other.’ (1994b: 405, fn.1) Similarly, for Horwich (DenS) is conceptually basic: our mastery of the concept <truth> is, roughly, to be explained in terms of our disposition to accept every instance of (DenS).

Leon Horsten’s (2009, 2011) inferentialist view, by contrast, explains <truth> in terms of its inferential role, which includes the introduction and elimination rules (T-In) and (T-Out):<sup>20</sup>

$$\begin{array}{ll} (\text{T-In}) & \frac{p.}{\text{'p' is true.}} \\ (\text{T-Out}) & \frac{\text{'p' is true.}}{p.} \end{array}$$

Again, substituting declarative sentences for 'p'.

The modest theory of Wolfgang Künne (2003: 336) analyses <truth> thus:

$$(\text{Mod}) \quad \forall x (x \text{ is true} \leftrightarrow \exists p (x = \langle p \rangle \ \& \ p)).$$

Where the angle-brackets, as they do here, serve as an operator that take a sentence to a singular term for the proposition that sentence voices (akin to the English ‘the proposition that’). Rather than ordinary objectual quantification or substitutional quantification, the second quantifier

<sup>19</sup> E.g., Field (1994a, 1994b), Horwich (1998). Horwich calls (DenS) the “equivalence schema”.

<sup>20</sup> The inferential role Horsten emphasises is one identified within a particular axiomatic theory of truth: Partial Kripke-Feferman (PKF). See Halbach & Horsten (2006). Horsten (2011: 39) is actually happy to allow that our *actual* concept of truth might not be that given in PKF; it is not uncommon for deflationists to maintain that their preferred account of truth is one we could “get by” with, while remaining neutral on our actual concept. I won’t go through this prescriptive spin on what I am describing as a descriptive project; it should be apparent enough how it goes.

here is supposed to express objectual quantification into sentential position. For Künne, our mastery of the concept <truth> is, in some sense, our mastery of this biconditional.

And, finally, prosententialists like Dorothy Grover, rather more radically, deny that ‘is true’ is really a predicate at all, and instead maintain that it is a prosentence-forming operator.<sup>21</sup> “Prosentences” are a kind of proform. Just as pronouns like ‘she’ “stand in” for nouns (‘Mary went straight to the restaurant; she was hungry’), proverbs like ‘do’ stand in for verbs (‘Dance as I do’), and proadjectives like ‘so’ stand in for adjectives (‘I’m sad, and want others to be so too’), so prosentences stand in for entire sentences. An example might be ‘so’ in the phrase: ‘Is it raining? If so, I’d rather stay here.’<sup>22</sup> According to Grover, ‘is true’ plays its expressive role by virtue of allowing us to form what are apparently prosentences, like ‘That is true’. That is, ‘That is true’ is a prosentence, which picks up the sentence that it “stands in” for anaphorically, like other proforms; so if you say ‘Ajax is a cat’ and I say ‘That is true’, this is a prosentence that stands in for the sentence ‘Ajax is a cat’.

Now, prosententialism to one side, it is worth re-emphasising that these “deflationary” analyses of the concept of truth are themselves compatible with ‘truth’ denoting a substantive property. However, setting out these different versions of deflationism, albeit briefly, renders salient the deflationary project. Any deflationary theory, if true, may be informative about the concept <truth> or our mastery of the use of the truth predicate. But they are uniformly utterly uninformative about *what truth is*. It is one thing to explain the meaning of ‘water’, or how users of that term are able to use it competently; it is something else altogether to say what *water* is. But that is all part of the deflationary program. According to the deflationist, *there simply isn’t anything interesting to say* about what truth is, beyond what there is to say about the concept or predicate. Deflationism is thus characterised by a distinctive *quietism* about the property of truth.

One might endorse this quietistic attitude on non-deflationary grounds, of course. One might be an alethic non-naturalist, for instance, and claim that truth is some *sui generis* property; or, perhaps even more bizarrely, think that truth is a fundamental property, along with charge, spin, or whatever. Neither is plausible, in my view, given a broadly scientific, naturalistic worldview. Alternatively, one might just harbour a global scepticism against the possibility of offering an informative philosophical account of any property. But then you’re not really a deflationist about truth, you’re just a philosophical pessimist (or a global deflationist of some kind).<sup>23</sup>

---

<sup>21</sup> See Grover, Kamp, & Belnap (1975), Grover (1992); see also Williams (1992) and Brandom (1994).

<sup>22</sup> Künne (2003: 68).

<sup>23</sup> One might endorse a kind of primitivism, which maintains that truth (property) is non-fundamental and yet, for some reason, we cannot say anything informative about what it consists in. (Compare Williamson (2000) on knowledge.) But if this is not to collapse into one of the options already described, one needs a principled reason to think we should be such primitivists about *truth in particular*. (E.g., Asay’s (2013: ch.5) re-imagining of Frege’s “treadmill” argument.) I am inclined to think of this as a “give up” position. In any case, if what I argue in the thesis is right, then it is wrong: we *can* say informative things about the nature of truth.

Setting these options aside, however, we wouldn't accept this kind of quietism when it comes to just any property.<sup>24</sup> So, if it is to be justified in the case of truth, then truth must be somehow anomalous. The deflationist's justification has two core aspects. The first is that 'truth', and the associated predicate 'is true', do not primarily function to denote or ascribe – or, generally, represent – some property in the world that might be of explanatory or normative import. Rather, the purpose of the truth predicate, the point of having it in our language at all, is to enhance its expressive capacity. In particular, it functions as a device for *endorsement* (and, in conjunction with negation, *rejection*; while 'false' is a device for *denial*),<sup>25</sup> allowing us to express endorsement of propositions without having to use sentences that voice those propositions.<sup>26</sup> The latter is especially useful when we cannot bring ourselves to use the relevant sentences, because we're too finite ('All the theorems of Peano Arithmetic are true!'), too ignorant ('I forget what it was, but I know that what she said was true!'), or just too lazy (the anaphoric 'That's true'). These uses aside, deflationists are typically keen to emphasise the odd quasi-redundancy of the truth predicate in truth-ascriptions like 'Grass is green' is true' and 'It is true that grass is green' – rhetorical flourish aside, these sentences don't seem to say much more nor less than 'Grass is green' does.<sup>27</sup>

The expressive capacity of truth is common ground. The distinctive deflationary claim is that this expressive capacity *exhausts* the utility of truth. We do not, for instance, need to postulate the property of truth to explain any phenomena. Rather, the only role for the truth predicate when it appears in explanations or norms is its expressive one, allowing us to formulate generalisations and so forth that would otherwise be difficult, cumbersome, or impossible to formulate. And, according to the deflationist, we can explain the truth predicate's expressive capacity (e.g., in one of ways outlined above) without having to think of 'truth' or 'is true' as denoting or ascribing any substantive or interesting property, about the nature of which there is anything interesting to say. As such, we do not have to think of truth as a substantive aspect of the world that our predicate or concept "tracks".<sup>28</sup> Indeed, early deflationists, like Ayer (1936: 87) and Strawson (1949: 84; 1950), denied

---

<sup>24</sup> For an explicit comparison, see the discussion of valence in Field (1972: 362-8). Of course, one might extend deflationism to *some* other properties, especially other semantic properties, like reference (indeed, one must, since if reference is substantive and truth reduces to reference, then one is an inflationist after all). The contention must be that these properties are anomalous in the same way.

<sup>25</sup> The point of distinguishing rejection from denial is to allow that one might rationally reject both a proposition and its negation (because neither is true, though neither is false), while denying a proposition commits one to accepting its negation, and vice versa.

<sup>26</sup> Or, perhaps, express endorsement of sentences without having to use those very sentences.

<sup>27</sup> The proto-deflationary *redundancy* theory of truth, associated with Ramsey (1927) and advocated by Ayer (1936: ch.5), maintained the literal equivalence in meaning between the two; a thesis that lives on in contemporary deflationism in e.g., Field's (1994a: 250-1) claim that "Grass is green' is true' is "cognitively equivalent" to 'Grass is green', relative to the assumption that 'Grass is green' exists. Similarly, Strawson's (1949, 1950) "performative" theory held that 'is true' is solely a device for performing various speech acts, like *confirming* or *admitting* or *conceding*. The central dissatisfaction with the redundancy and performative theories was the inability to *explain* the expressive capacity of the truth predicate, which contemporary deflationary theories endeavour to do.

<sup>28</sup> To be clear, I take this kind of justification for the associated quietism to be what's constitutive of deflationism.



that truth is a property at all; but since it determines a function from possible worlds to extensions, contemporary deflationists are typically happy to allow that ‘truth’ denotes a property, but just not one with any substantive or interesting nature.<sup>29</sup> While *instances* of a relevant schema, like (DisS) or (DenS), may be of explanatory or normative import, truth *itself*, the story goes, is explanatorily and normatively impotent. It is thus the deflationist’s anti-explanatory contention that is the focal point of the debate between inflationists and deflationists, and which will therefore be the focus of Chapter 3.<sup>30, 31</sup>

### 1.2.2 Realist Inflationism

Inflationary theories deny that semantic properties like truth are anomalous in the way that deflationists contend that they are. Inflationists differ on what the nature of the property of truth consists in, however. For our purposes, we will divide inflationary theories into two families: “realist” and “anti-realist”.

Probably the most prominent rival to deflationism is the correspondence theory of truth. Correspondence theorists hold that truth is a relational property, which consists in the truthbearer standing in a substantive relation – the correspondence relation – to something that we will call (following the literature) a *truthmaker*, which is roughly that part of reality that the truthbearer is about.<sup>32</sup> This is schematic: a particular correspondence theory will, in addition to an account of truthbearers, need to explicate the nature of the correspondence relation and the truthmakers.

There are two prominent varieties here. On “fact-based” theories, each true truthbearer *as a whole* stands in the correspondence relation to a *fact* (or fact-like entity, e.g., a state of affairs). For instance, ‘Grass is green’ will correspond to the fact that grass is green. The fact that grass is green will, presumably, be grounded in there being a property, greenness, that is exemplified by grass. “Object-based” theories, on the other hand, have it that the *parts* of a truthbearer – subpropositional or subsentential elements, for instance (concepts or words) – stand in the correspondence relation to parts of reality; the truth of a truthbearer as a whole is then to be analysed in terms of the correspondence of its parts. For instance, ‘grass’ will correspond to grass, ‘green’ will correspond to greenness, and ‘Grass is green’ will be true just in case the object that ‘grass’ corresponds to exemplifies the property that ‘green’ corresponds to; i.e., iff grass is green. We might thus see object-

---

<sup>29</sup> Thus with Edwards (2013b) and Asay (2014), we can think of the deflationism as the view that truth is merely an abundant property, rather than a (comparatively) sparse one.

<sup>30</sup> I am thus sympathetic with Eklund’s (2010) suggestion that the dialectic would benefit if would-be deflationists stopped with the claim that the sole point of the truth predicate is its expressive utility, rather than going on to offer distinctive – especially schematic – conceptual analyses; a position he calls “sophisticated rejectionism”. (Eklund himself is no rejectionist.)

<sup>31</sup> Note that, while I assume that semantics is truth-conditional, which is often taken to be inconsistent with deflationism about truth, I reject deflationism on quite independent grounds, so this shouldn’t be dialectically problematic.

<sup>32</sup> In using the term, I do not commit myself to any of the claims of so-called truthmaker theory.

based correspondence theories as reducing truth to denotation and offering a correspondence theory of that.<sup>33</sup>

Either way, correspondence theories are typically taken to wear their realist credentials on their sleeve. For on these accounts, it is part of the very nature of truth that the part of reality that the true truthbearer talks about *exists*: if ‘Grass is green’ is true then there must be such thing as grass and such thing as greenness, for this is constitutive of the very nature of truth.<sup>34</sup> This is amenable to the realist view that intentional entities are true in virtue of accurately representing a world that exists independently of us. We should be careful, however: for the correspondence theory in and of itself says nothing about the *nature* of (those things that constitute) the truthmakers<sup>35</sup> other than that they are capable of standing in the correspondence relation, and until we have an explication of *that* there are no constraints on what truthmakers might be. They may, for instance, be paradigmatic mental entities like *ideas*; indeed, McTaggart, it seems, was a correspondence theorist and an idealist (McDaniel 2016: §5). Nonetheless, because of its structural amenability to realism, we shall label correspondence theories “realist” inflationism.<sup>36</sup>

Lynch (2009: ch.2) suggests that we might see modern causal or teleosemantic theories of *representation* as inheritors of the correspondence theory of truth.<sup>37</sup> In Lynch’s hands, a causal view of representation explains why a concept <F> represents Fs (or F-ness) in terms of some causal relation that holds between <F> and Fs,<sup>38</sup> while Lynch’s teleosemanticist explains representation in terms of biological function: <F> represents Fs because natural selection has chosen it to, say, be tokened in the presence of Fs.<sup>39</sup> The object-based correspondence theorist can see these theories as offering accounts of denotation; the thought, then, is roughly that we can conjoin our causal or

---

<sup>33</sup> Note that, if *Russellian* propositions are the primary truthbearers, then we end up very close to saying that propositions are true in virtue of corresponding to themselves. This looks much closer to an *identity* theory of truth, which says that true propositions are *identical to* facts. (E.g., Hornsby 1997, Dodd 2000.) (It doesn’t quite get us all the way there, since whatever is supposed to hold together a Russellian proposition won’t be, e.g., the object exemplifying the property. Whether or not the identity theory itself counts as a correspondence theory is a terminological issue that turns on whether or not we legislate that the correspondence relation must hold between distinct entities.) Note that the argument in favour of the correspondence theory in Ch.3 and Ch.4 is an argument in favour of there being a substantive relation between certain true beliefs and the part of the world they are about. Now, if the objects of belief are Russellian propositions, then this does not rule in favour of a correspondence theory, rather than the identity theory *for propositions*; but we still need a correspondence theory of truth *for the beliefs*. At which point, the propositions are no longer primary. (This is one reason why the nature of truthbearers matters here.) For this reason, I set the identity theory aside from hereon. For what it’s worth, I am not attracted to the identity theory of truth nor the (naïve) Russellian view of propositions for quite independent reasons.

<sup>34</sup> I do not mean to imply that this commits us to any particularly substantive views about the nature of, e.g., properties.

<sup>35</sup> Moving forward, I’ll take the bracketed phrase as implicit.

<sup>36</sup> For good introductory overviews of the correspondence theory, see Künne (2003: ch.3) and David (2016).

<sup>37</sup> Indeed, he describes them as ‘the most plausible successor to the traditional correspondence theory.’ (2009: 22)

<sup>38</sup> See Field (1972), Stampe (1977), Devitt (1981, 1987), Evans (1982), Fodor (1987, 1990, 1994), Lloyd (1987), Rupert (1999), Ryder (2004).

<sup>39</sup> For sophisticated teleosemantic views, see especially Dretske (1981, 1988, 1995), Millikan (1984), and Papineau (1987); for an excellent overview, see Neander (2012). Note that, contra Lynch, some versions of these views (especially Papineau’s) are probably more amenable to a fact-based correspondence theory than an object-based one.

teleosemantic account of representation with the existence of the relevant truthmaker to obtain our correspondence theory of truth. This is an intriguing and substantive way of unpacking the correspondence theory, and it will be useful throughout to have as a concrete example. (Correspondence theories are often accused of being platitudinous and hollow.)

(Note, however, that Lynch’s theory requires that concepts be both (i) able to enter into causal interactions, and (ii) the components of primary truthbearers. He is pretty clear at this point that he is taking concepts to be mental particulars, so if we’re thinking of propositions as composed of concepts, then we might have to subscribe to something like a Soamesian view of propositions. Alternatively, we might think of concepts as Mentalese words, and take the primary truthbearers to be Mentalese sentences. For those who find these options unattractive, we should note that one can endorse causal or teleosemantic accounts of representation without thinking of them as offering an account of truth, but rather an account of *which* proposition is the content of a particular mental state (perhaps by virtue of explaining the content of particular mental entities). One would then need a distinct account of truth for the proposition. This is significant, insofar as Lynch’s own primary argument for pluralism relies on his broadly causal interpretation of the correspondence theory (see §2.2.3).)

### 1.2.3 *Anti-Realist Inflationism*

For terminological ease, I will group together all those inflationary theories that do not analyse truth in terms of the truthbearer standing in a relation to the part of reality that the truthbearer is about “anti-realist” theories. (See Küne (2003: 21) for a list of biconditional analyses of this form extracted from *inter alia* William James, Charles Peirce, Nelson Goodman, Hilary Putnam, Michael Dummett, and Crispin Wright.)

The most prominent families of anti-realist theories in this sense are “coherence” theories and “pragmatist” theories. In outline, the coherence theory says that the truth of a primary truthbearer consists in the belief that p being coherent with some specified body of beliefs B. This need not entail that beliefs are the primary truthbearers; propositions may be the primary truthbearers, but the truth of the proposition that p nonetheless consists in the belief that p being coherent. Different coherence theories may differ over the nature of coherence or which beliefs appear in B.<sup>40</sup>

It is difficult to identify an underlying thesis uniting pragmatist theories, but there are two prominent stereotypes. The first, largely inspired by the work of William James (1907: ch.6), characterises truth in terms of utility: what is true is what is in some sense useful to believe. The second, largely inspired by Charles Peirce (1878), characterises truth in terms of what all investigators

---

<sup>40</sup> For a useful overview of coherence theories, see Walker (2001).

would come to believe if investigation into the question of whether or not  $p$  were pursued long enough – at the fabled “end of inquiry”.<sup>41</sup>

Also among the anti-realist theories, and most important for our purposes, are “epistemic” theories of truth, which bear some similarity to both their coherence and pragmatist forebears. Broadly speaking, epistemic theories aim to analyse truth in terms of some epistemic property, like verification or warrant. Dummett’s (e.g., 1978, 1993) verificationism and “semantic” anti-realism is prominent in this tradition. It will be useful for our purposes to describe two such theories, to use as reference points. The first is *superwarrant*, a development of Wright’s notion of *superassertibility* (e.g., 1993: ch.14, §3; 1992: ch.2; the below definition is primarily based on the explicit formulation in 1996: 865):

#### *Superwarrant*

For any  $p$ ,  $p$  is true iff there exists some state of information  $I$  that the world, constituted as it is, would generate in a suitably receptive, investigating subject, such that the belief that  $p$  is warranted in  $I$  and remains warranted in  $I$  no matter how  $I$  is enlarged or improved upon.

Thus to be true is, on this account, to have a particularly stable warrant, such that this warrant would survive arbitrary improvement to the state of information that provides the warrant. The paradigm here is supposed to be mathematical proof: Andrew Wiles’ proof of Fermat’s Last Theorem would provide a warrant for believing it for anyone able to go through the relevant steps to understand it (and thus access the relevant state of information), and this warrant (given that the proof contains no errors) will survive arbitrary improvement to and enlargement upon that state of information.<sup>42</sup> I follow Lynch (2009: ch.2) in reformulating the principle in terms of warrant for belief rather than assertion.

Superwarrant will be our touchstone anti-realist theory of truth moving forward, but I’d like to also mention Lynch’s preferred such theory, which he calls *concordance*, and which is formulated for moral propositions in particular:

#### *Concordance*

For any moral  $p$ ,  $p$  is true iff there exists a set of beliefs  $F$  such that: (i) on balance, adding  $p$  to  $F$  would make  $F$  more coherent; (ii) the first condition would continue to hold through any successive and additional improvements to  $F$ ; and (iii) (the content of)  $F$ ’s non-moral beliefs correspond to reality.<sup>43</sup>

---

<sup>41</sup> Peirce’s view might be better characterised as an epistemic theory.

<sup>42</sup> ‘*Superassertibility* – the generalization of mathematical provability we seek’ (Wright 1993: 414).

<sup>43</sup> See Lynch (2009: ch.8; 2013a) as well as Capps et al. (2008). Recently Lynch has added a further condition to concordance: that *the proposition that (i) and (ii) hold of  $p$*  – call this proposition  $Q$  – is not true in virtue of (i) and (ii) holding of  $Q$ . But this just seems to follow from the fact that this is not a moral proposition, rather than being part of the very nature of the property of concordance. Lynch does not actually mention

I do not mention this because I take it to be a particularly plausible theory, but because condition (iii) is illustratively useful: if one is willing to be a *pluralist* about truth, in a sense that will become clear in the following section, then one can analyse one truth property *in terms of* another. The epistemic theory I develop in Chapter 5 – *weak super-stability* – is broadly an inheritor of superassertibility, superwarrant, and concordance.

Two features of anti-realist theories worth taking note of. First, suppose that <eating meat is wrong> is true in virtue of being superwarranted. This does not entail that there is no property of wrongness that eating meat exemplifies. On the contrary, the truth of the proposition entails that eating meat is within the extension of ‘wrong’, so there will be such a property in at least a minimal sense. What the anti-realist denies is a role for the relevant objects and properties *in the analysis of the nature of the truth of the proposition*; but this does not (by itself, anyway) entail that there are no such objects and properties. This, we will see, is analogous to expressivist’s position in the theory of meaning: the metaethical expressivist denies moral properties a role in the explanation of the *meaning* of the proposition, but that does not entail that there are no such properties (§4.2).

Second, anti-realist theories share a commitment to the *in-principle accessibility* of truth, in the sense that any truth that is comprehensible to human beings can become the content of a justified human belief.<sup>44</sup> This is what it means (at least, in my mouth) to say that a theory of truth is *epistemically constrained*, or that truth is not *evident-transcendent*, or does not outrun *rational acceptability*. Indeed, this commitment is historically central to the anti-realist aspirations of such theories: the truth is somehow limited by the limits of human comprehension.

## 1.3 Pluralisms

### 1.3.1 <Truth>: Pluralism, Monism, Inconsistency

‘Pluralism’, much like ‘deflationism’ and ‘inflationism’, picks out a family of theories. This section aims to provide the most complete taxonomy of truth pluralism to date. Many of the possible views here are, I think, systematically overlooked. As well as intrinsic interest, it matters both dialectically and methodologically that we are alive to the theoretical options. In this first subsection, I introduce conceptual pluralism about truth, and its important rivals; the second subsection concerns property pluralism.

Conceptual pluralists about truth think that there is more than one truth concept, i.e., that ‘is true’ voices more than one concept, and is thus ambiguous (even when applied to one kind of

---

correspondence in condition (iii), but rather just mentions truth; to explain the reasons why would at this stage mean getting ahead of ourselves.

<sup>44</sup> Künne (2003: 20).

truthbearer).<sup>45</sup> The clearest example here is Max Kölbel (2008, 2013).<sup>46</sup> Kölbel’s argument relies on the intuition that certain truthbearers – those concerning matters of taste, like ‘Skiing is more fun than swimming’ or ‘Paul Foot is funny’ – *cannot be either true or false*. Nonetheless, Kölbel argues, we can still use ‘is true’ as a device for endorsing such claims: if you say ‘Paul Foot is funny’ and I agree, I might say, ‘That’s true.’ Kölbel even provides some very limited empirical data that ordinary speakers are inclined to make both of these judgements, since his students replied to questions on a questionnaire by using ‘true’ (or ‘false’) to (dis)agree with a certain statement concerning the funniness of a particular comedian, *and* on the same questionnaire endorsed the claim that such claims can be neither true nor false. The best explanation, suggests Kölbel, is that ‘is true’ is ambiguous. It has a deflationary meaning, <truth<sub>d</sub>>, which we voice when we use it as a device for endorsement, and which can be applied to any claim. But it also has a more substantive meaning, <truth<sub>s</sub>>, which comes saddled with a notion of *objectivity*: to say that something is true<sub>s</sub> is to say it is both true<sub>d</sub> and objective, more or less. It is this notion of truth that we refuse to ascribe in matters of taste.

What is important to note about this suggestion is that this is no radical ambiguity; rather, being true<sub>s</sub> is a *way* of being true<sub>d</sub>. Kölbel even suggests that our language contains other such ambiguities. Some usages of ‘duck’, for instance, seem insensitive to sex; but sometimes we mean <female duck>, as when we say, ‘That’s not a duck; that’s a drake!’ A similar phenomenon appears with ‘dog’, which can, it seems, mean <male dog> in particular.

Most pluralists, however, are pluralists about the property and explicitly *monists* with regards to the concept; this goes for both of the most prominent pluralists in the recent tradition, Crispin Wright and Michael Lynch. Indeed, both Wright and Lynch (following him) pursue a *platitudo-based* approach to conceptual analysis, which involves assembling a number of apparently truistic or platitudinous claims about truth, such that were someone to deny them, we would immediately doubt that they meant the same thing by ‘truth’ as we do.<sup>47</sup> Wright’s platitudes, for instance, include:

- ‘that to assert is to present as true;
- that any truth-apt content has a significant negation which is likewise truth-apt;
- that to be true is to correspond to the facts;
- that a statement may be justified without being true, and vice versa;<sup>48</sup>

---

<sup>45</sup> Or, similarly, one might maintain that truth predicates in *different* languages voice different concepts; I have seen this view defended by Jeremy Wyatt, who suggested (based on work by Wiredu (2004)) that the Akan ‘ete sa’ voices a different concept to the English ‘is true’. The position is awkward, though perhaps not untenable, because we can ask why they both count as *truth* predicates (instances of the same kind; inter-translatable) if they voice different concepts, and hence mean different things. One needs a principled account of what a truth predicate is, but which *prima facie* can’t appeal to a unified concept.

<sup>46</sup> For related views, see Field (1994a), McGee & McLaughlin (1995), and McGee (2005).

<sup>47</sup> This methodology is prominently advocated by Jackson (1998) and is associated with the “Canberra Plan”.

<sup>48</sup> Wright (1992: 34).

while Lynch’s favourite triumvirate is:

*Objectivity*: The belief that p is true if, and only if, with respect to the belief that p, things are as they are believed to be...

*Norm of Belief*: It is *prima facie* correct to believe that p if, and only if, the proposition that p is true...

*End of Inquiry*: Other things being equal, true beliefs are a worthy goal of inquiry.<sup>49</sup>

They contend that this analysis identifies the unified concept voiced by ‘truth’. But, they continue, this still allows for property pluralism because these platitudes might be “satisfied” by different properties for different (sets of) truthbearers. Perhaps some propositions are *prima facie* correct to believe when they correspond with the facts, but others are *prima facie* correct to believe when they are superwarranted. Note that this requires, on the face of it, that our platitudes not be universally quantified, for then there’s no such thing as satisfying the platitudes for a restricted set of truthbearers.<sup>50</sup> The precise concept/property relationship here thus needs finessing. Lynch, for instance, takes his platitudes (‘truisms’, in his terms) to identify a unified *functional role*: truth is then a unified property that something has iff it possesses some property that plays this role, but the role itself is multiply realisable;<sup>51</sup> Lynch is thus an inflationary moderate pluralist, in the terms we’ll introduce below.<sup>52</sup> Wright seems to deny that there is a unified truth property, but he is less clear on what exactly he takes the relationship to be here.<sup>53</sup>

A final view on the concept of truth that we need to have on the table is the *inconsistency theory*. An explanation of the semantic paradoxes involving truth (especially the liar paradox and the Curry paradox) that has been gaining traction recently is that ‘is true’ voices an inconsistent concept.<sup>54</sup> The way this is usually cashed out is that there are rules that are in some sense constitutive of the meaning of the word (and thus the concept) that give rise (perhaps in conjunction with empirical facts) to inconsistency. For instance, (T-In) and (T-Out) allow us, with minimal further resources, to derive the liar and Curry paradoxes, so if we take them to be constitutive of <truth>, then <truth> may well be inconsistent. In the final chapter, I will suggest a novel solution of the liar paradox that

---

<sup>49</sup> Lynch (2009: 8, 10, 12).

<sup>50</sup> As pointed out by Edwards (2011: 34-9).

<sup>51</sup> For excellent discussion, including a very helpful “tidying-up” of Lynch’s view, see David (2013).

<sup>52</sup> More on Lynch’s functionalism later (§6.1.3).

<sup>53</sup> In recent writings, Wright (2013) seems to favour the analogy between truth and winning made famous by Dummett (1959) (as does Edwards (2011, 2013c, fc)). Just as what it takes to win can be very different from game to game (crossing a line first, scoring the most points, scoring the least points, potting a ball first), so what it takes to be true might vary from truthbearer to truthbearer. My central worry about this analogy is that the one/many relation in the case of winning looks like something that is crying out for explanation, rather than something we antecedently understand and can use to shed light on truth. The best it gives us is arguably a *licence for optimism*: since some one/many relation holds in the case of winning, presumably the same such relation might hold for truth. But this still leaves everything to be explained.

<sup>54</sup> Eklund (2002a, 2002b, 2008), Azzouni (2006, 2007), Patterson (2007a, 2007b, 2009), Scharp (2007a, 2007b, 2013).

combines an inconsistency theory of the concept with the inflationary pluralism that is advocated in the prior chapters.

What implications does the inconsistency theory have for the property of truth? This depends on one's semantics for inconsistent concepts. It might, at first blush, seem to recommend a kind of nihilism about truth: we do not want to say that the world contains an inconsistent property, so perhaps there is no property denoted. Matti Eklund, however, suggests an indeterminacy view: the concept denotes whichever property makes its constitutive principles 'come out as nearly true as possible' (2002b: 269), and it is indeterminate, with 'truth', what this property is. Kevin Scharp (2013: ch.9), by contrast, suggests an assessment-sensitive view, such that the extension of 'true' varies between what is denoted by his "replacement" concepts, <ascending-truth> and <descending-truth>.<sup>55</sup> I return to this issue at §8.1.2. I will, until this point, endeavour to be neutral about the nature of <truth>.

### 1.3.2 A Panoply of Property Pluralisms

The core thesis of pluralism about the *property* of truth is that there is more than one truth property; not just, of course, that more than one of the properties outlined above exist, but that the truth of some truthbearers  $S_1$  consists in one property, and the truth of some truthbearers  $S_2$  consists in another. We can straightaway outline four important parameters for variation given this core thesis: the inflationary/deflationary status of the properties; the nature of the inflationary properties endorsed; the application of the truth properties; and whether or not they acknowledge an unrestricted truth property.

These parameters are to a large extent orthogonal, so the taxonomy here quickly becomes highly complex. The reader is not expected to follow and remember all of various views detailed here at once; I will outline the relevant views again when they come up in discussion. Rather, this subsection can function as a detailed reference point for the reader to refer back to. (In case the reader would find it useful, I have provided a flowchart of many of the views in *Figure 1*, with the positions I'll defend in the ensuing chapters highlighted in bold.) Most important for my purposes in the thesis is the view I call *domain-restricted anti-/realist inflationary pluralism*: the view that truth consists in a realist truth property for some truthbearers  $S_1$  and consists in an anti-realist truth property for a *disjoint* set of truthbearers  $S_2$ ; and *which* set a particular truthbearer falls into is determined by its subject matter or "domain", e.g., whether it's about physics, mathematics, or ethics. With that emphasised, let's turn to the full taxonomy.

First, one might be a purely deflationary pluralist. This would require holding that there is a multiplicity of truth predicates – 'T<sub>1</sub>', 'T<sub>2</sub>', etc. – such that the analogues of (T-In) and (T-Out) are

---

<sup>55</sup> <Ascending-truth> has the analogue of (T-In) as a constitutive principle; <descending-truth> the analogue of (T-Out).



valid for different sets of truthbearers, but that *none* of these predicates designates a substantive property. For instance, on a deflationary interpretation of Tarski's view (discussed a greater length in Chapter 8), we have language-relative truth predicates. Another view that is sometimes cited as an example is Field (2008), who defines a *determinateness* operator that can be iterated infinitely, such that something can fail to be determinately true without failing to be true; can fail to be determinately determinately true, without failing to be determinately true; can fail to be determinately determinately determinately true, without failing to be determinately determinately true; and so on.<sup>56</sup> Each truth predicate so formulated has different application, and is deflationary. Beall (2013) also, and explicitly, outlines a deflationary pluralism.

A *hybrid* pluralist endorses both inflationary and deflationary truth properties (at least one of each), and of particular import is the species of this genus I will call the *partial inflationist*: the partial inflationist holds that some set of truthbearers is *only* apt for a deflationary truth property  $T_D$ , while others are (perhaps also)<sup>57</sup> apt for an inflationary property  $T_I$ .<sup>58</sup> Consider, for instance, Kölbel's conceptual pluralism applied at the level of properties: any declarative sentence, according to Kölbel, can be true in a purely deflationary sense, but only certain sentences can be true in a more robust inflationary sense. Thus a proper subset of those sentences apt for  $T_D$  are apt for  $T_I$ .<sup>59</sup> Hybrid pluralists that are *not* partial inflationists will hold that, while *every* truthbearer will be apt for an inflationary truth property of some kind or other, there are also deflationary truth properties for which at least some truthbearers are *also* apt. For instance, we might suppose that the domain of truthbearers is partitioned into two sets,  $D_1$  and  $D_2$ , such that those in  $D_1$  are apt for inflationary property  $T_1$  and those in  $D_2$  are apt for inflationary property  $T_2$ ; but that *all* truthbearers are apt for some deflationary truth property  $T_D$ . Hybrid pluralists, including partial inflationists, might themselves endorse more than one deflationary property, in which case they are also deflationary pluralists; more than one inflationary property, and be inflationary pluralists (of any variety); or both.

A purely inflationary pluralist will (unsurprisingly) endorse a multiplicity of inflationary truth properties. Here we introduce our second parameter of variation: different inflationary pluralists will endorse importantly different *kinds* of truth property. Realist inflationary pluralists – often called “correspondence pluralists” – endorse a variety of different correspondence relations that hold

---

<sup>56</sup> Perhaps it's not quite right to think of the truth predicate modified by the determinateness operator (however many times) as itself a distinct truth predicate. Still, I mention Field here because he is sometimes cited as an example of a deflationary pluralist (e.g., Cotnoir (2013b: 563)) and it's a useful example.

<sup>57</sup> On this, see the discussion of intersectional vs. disjoint pluralism below.

<sup>58</sup> Wright's (1992) pluralism sometimes sounds like partial inflationism, since he talks of “heavyweight” truth in some domains, and “lightweight” truth in others. This is compounded by the fact that Wright's anti-deflationary argument is that deflationists cannot make sense of the fact that the norm of truth can outrun that of warranted assertibility; and, of course, *in a sense* superassertibility doesn't outrun warranted assertibility, so Wright might think superassertibility is compatible with deflationism. But I think it is better to view Wright's proposal as an (anti-/realist) inflationary pluralism. I have seen Filippo Ferrari & Sebastiano Moruzzi argue in favour of partial inflationism, precisely by resisting Wright's anti-deflationary argument for certain truthbearers (those concerning “basic taste”). See also Wright (1998b, 2001).

<sup>59</sup> Kölbel defines his “substantive” truth property in terms of objectivity, so what I am calling an “inflationary” truth property here does not look a lot like those we encountered in the previous section.

between truthbearers and the world. For instance, Gila Sher (1998, 2004, 2005, 2013, 2015) holds that there are different “forms” of correspondence: where in physics ‘reference... is largely based on a direct *causal* relation between physical expressions and objects in the world’ (2013: 162), mathematical and logical truths somehow correspond to the formal structure of reality. For Barnard & Horgan (2006, 2013), truths concerning the fundamental level of reality “directly” correspond, while all other truths only “indirectly” correspond.<sup>60</sup> Anti-realist inflationary pluralists would endorse a variety of anti-realist truth properties. Perhaps this view has been defended, but nothing comes to mind.

Of primary interest in this thesis will be another hybrid view, which I will call *anti-/realist inflationary pluralism* (or just anti-/realist pluralism), which endorses both realist and anti-realist truth properties, at least one of each. Before giving examples, we must lay out our third parameter, orthogonal to the previous two, which concerns *application* of the truth properties – that is, *which* truthbearers the different properties are relevant for. According to *disjoint* pluralism, disjoint sets of truthbearers are apt for distinct truth properties:  $T_1$  is relevant for truthbearers in  $S_1$ ,  $T_2$  is relevant for truthbearers in  $S_2$ , and there is no  $p$  such that  $p \in S_1$  and  $p \in S_2$ . *Intersectional* pluralism, by contrast, would hold that the application of the truth properties intersects, either entirely or in part.<sup>61</sup> Intersectional pluralism is unappealing for the inflationary pluralist, since it seems possible in general that a truthbearer may possess  $T_1$  but not  $T_2$ . However, it is less immediately problematic if there is the right kind of metaphysical relationship between the truth properties. Consider, again, a Kölbelian partial inflationism: while all truthbearers are apt for  $\text{true}_d$ , certain truthbearers are *also* apt for  $\text{true}_s$ , but there is no possibility of them being  $\text{true}_s$  without being  $\text{true}_d$ , since being  $\text{true}_s$  is a *way* of being  $\text{true}_d$ ; just as there is no way of being a drake without being a duck.<sup>62</sup> (This will also be relevant when it comes to moderate pluralism below.) Hybrid options are available here too: if one endorses three truth properties, say, then  $T_1$  and  $T_2$  may intersect, while  $T_3$  is disjoint from both.

However, in its most prominent form, inflationary pluralism is disjoint. The disjoint pluralist takes on the burden of saying which truthbearers fall into which set: of individuating the different sets of truthbearers for which the different properties are relevant. The most familiar way of doing so is via *domain*, that is, roughly, the kind of thing that the truthbearer is about. Consider the following list:

- |            |                                 |
|------------|---------------------------------|
| (Ordinary) | Ajax is a cat.                  |
| (Maths)    | $7 > 3$ .                       |
| (Morality) | The Invasion of Iraq was wrong. |

<sup>60</sup> In my view, it’s not obvious that Barnard & Horgan’s “indirect” correspondence is really a correspondence theory of truth.

<sup>61</sup> Intersectional pluralism is mentioned by Wright (2005: 7).

<sup>62</sup> To extinguish the worry entirely, we also need to argue that none of *these* truthbearers can be  $\text{true}_d$  without being  $\text{true}_s$ .

(Art)	<i>Deadwing</i> is beautiful.
(Theory)	Electrons have negative charge.
(Social)	The Iraq War was expensive.
(Mind)	Jim is happy.
(Religion)	God is love.
(Fiction)	Lyra Belacqua can read the alethiometer.
(Comedy)	Paul Foot is funny.
(Law)	The Invasion of Iraq was illegal.

Intuitively, these each concern importantly different subject matters: mathematics, morality, aesthetics, religion, physics, and so on; they fall into different “domains”. For the *domain-restricted* pluralist, which truth property is relevant for a particular truthbearer depends on which domain it falls into. Of course, these intuitive differences are merely a starting point: it is incumbent on the pluralist to say exactly what it is about different domains that are apt for different properties that makes them so. (This topic will be prominent in the next chapter and Chapter 7.) To be explicit: it is no part of the domain-restricted pluralist’s view that *each* domain has its own truth property, but rather that disjoint sets of truthbearers are apt for different truth properties, and which set a particular truthbearer falls into is *determined by* its domain. For instance, moral and aesthetic truthbearers, qua moral and aesthetic, may fall into one set, while physical and mental truthbearers, qua physical and mental, fall into another; the different sets may thus incorporate multiple domains. However, at certain points it will be useful to use ‘domain’ to refer to the disjoint sets themselves, rather than domains strictly speaking contained therein. I trust no confusion will result.

The anti-/realist domain-restricted pluralist holds that truthbearers in one domain are apt for a realist truth property, while truthbearers in another are apt for an anti-realist truth property. The application of each of the different truth properties is thus “restricted” to particular (sets of) domains, hence the name. Lynch, for example, holds that propositions about ordinary macroscopic objects like <the cat is on the mat> are apt for correspondence, but that moral propositions are apt for concordance. Views of this kind dominate pluralist theorising: coherence truth properties are sometimes suggested for the mathematical domain; epistemic truth properties for normative domains, like the ethical and aesthetic; and correspondence typically for truthbearers concerning ordinary “medium-sized dry goods” and “scientific” domains like physics and chemistry. The kinds of motivation given for this view are thus, unsurprisingly, broadly along the lines of localised realism and anti-realism: arguments in favour of realism in some domains and anti-realism in others. As we shall see, however, it is no easy feat to argue from these kinds of considerations to domain-restricted anti-/realist pluralism.

Anti-/realist pluralists can also endorse more than one realist truth property, or more than one anti-realist truth property. Note also that domain-restricted pluralism is only a species of the

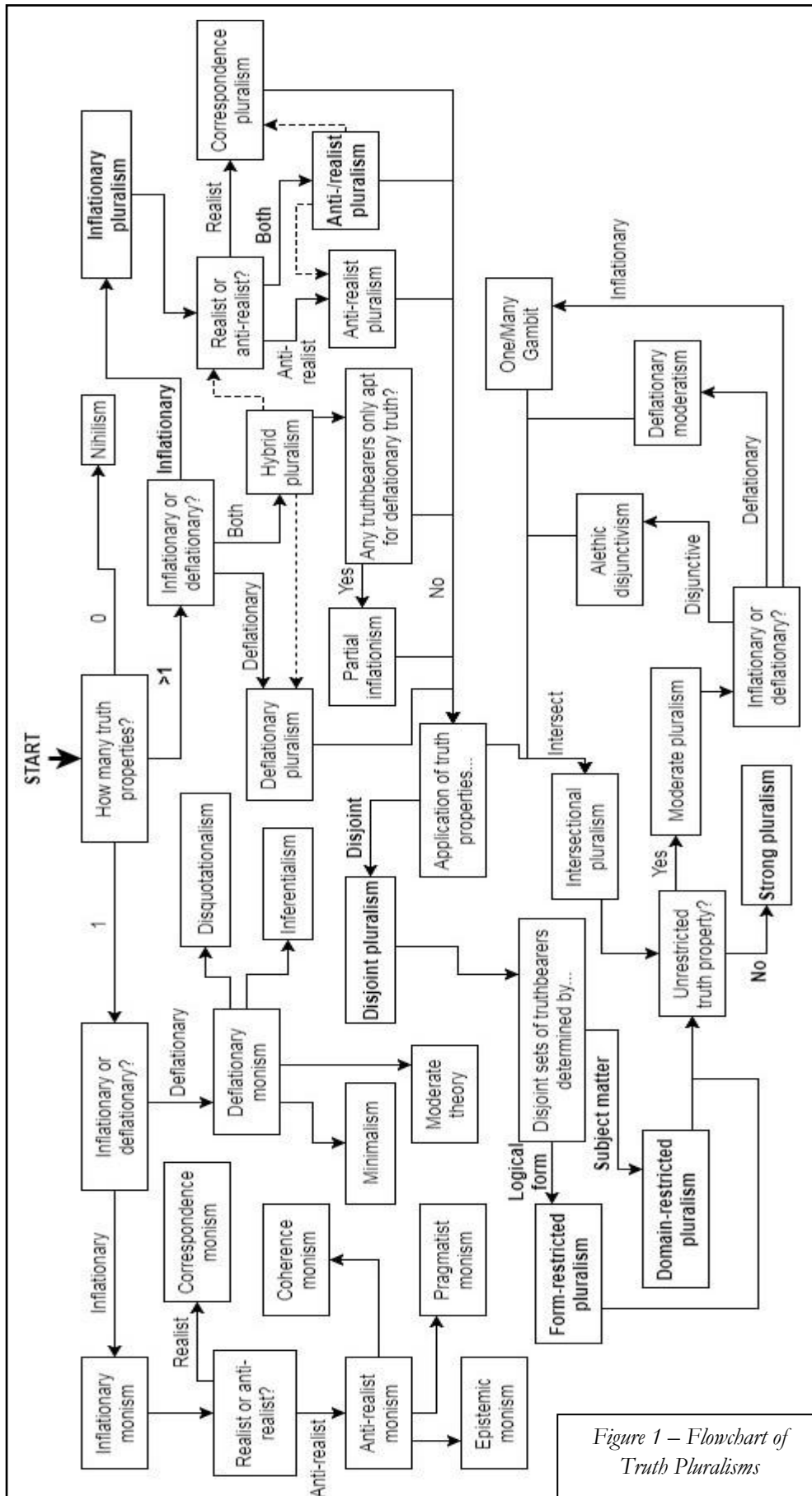


Figure 1 – Flowchart of Truth Pluralisms

disjoint pluralism genus. As such, an argument in favour of disjoint pluralism is not *ipso facto* an argument for domain-restricted pluralism in particular. This will be important in §2.2.3.

We shall encounter another variety of disjoint pluralism in Part III: *form-restricted* pluralism, which holds that the relevant truth property for a truthbearer depends on its logical form.

One last parameter that it's important to have on the table. So-called *moderate* property pluralists also concede the existence of a truth property with *unrestricted* application,  $T_U$ . (Moderate pluralists are *ipso facto* intersectional pluralists.) This itself comes in two importantly different varieties: deflationary moderates think that  $T_U$  is a deflationary property (and are thus hybrid pluralists); inflationary moderates think  $T_U$  is inflationary. This distinction is vague, because the inflationary/deflationary distinction is vague, but there are clear cases on either side of the boundary. Most prominent in the literature are inflationary moderates.

What is striking about inflationary moderates is that they concede the central contention of inflationary *monism* about truth: that there is a substantive unified truth property in common among all the truths. In order to retain their inflationary pluralism in the wake of this monistic concession, inflationary moderates endorse what I call the *One/Many Gambit*, which holds that there is a one/many relation, such as genus/species, determinable/determinate, functional role/realiser, that holds between  $T_U$  and the restricted truth properties, such that *being  $T_1$  is a way of being  $T_U$* . Such a view is thought to be a necessary amendment to domain-restricted inflationary pluralism; I criticise this contention in Part III, especially §6.1.3 – see there for citations.

A deflationary moderate pluralist would hold that, while different sets of truthbearers are apt for different inflationary truth properties, there is nonetheless a deflationary truth property that applies to all truthbearers. A sort of in-between view here is what has been called *alethic disjunctivism*.<sup>63</sup> Suppose that one is a domain-restricted inflationary pluralist:  $T_1$  is relevant for  $D_1$ ,  $T_2$  is relevant for  $D_2$ , ..., and  $T_n$  is relevant for  $D_n$ . It then seems we can define a unified truth property thus:<sup>64</sup>

*Disjunctive- $T_U$*

$$\forall p (T_U(p) \leftrightarrow ((p \in D_1 \ \& \ T_1(p)) \vee (p \in D_2 \ \& \ T_2(p)) \vee \dots \vee (p \in D_n \ \& \ T_n(p))))$$

While “unified”, this property is not grounded in any genuine similarities between the properties disjoined on the right-hand side, so the property appears to be merely abundant, and thus somewhat deflationary. It is not as deflationary as the traditional deflationist contends; there are, after all, genuine similarities among significant sets of truths. But the unrestricted property is not

<sup>63</sup> Pedersen (2006, 2010), Edwards (2012b), Pedersen & Wright (2013b).

<sup>64</sup> This received some attention (especially in Pedersen (2010) and Wright (2012)) as a potential worry for pluralism, which seems to hereby “collapse” into monism. But there is no real worry here; this is no *inflationary* monism.

the type of property that can be invoked in explanations, as anything other than an expressive device.<sup>65</sup>

Moderate pluralists are contrasted with *strong* pluralists, who deny the existence of *any* such unified property. In the final chapter, I will suggest that this position is motivated by the liar paradox, since the existence of an unrestricted truth property allows us to formulate a sentence that says of itself that it does not possess that very property; and I show how my form-restricted pluralism allows us to avoid paradox.

Any property pluralism gives rise to a kind of conceptual pluralism, insofar as each of the properties can be picked out by a different concept. More interesting, however, are *descriptive* analogues of property pluralism at the level of concept, i.e., ambiguity views that hold that the predicate ‘is true’, as it appears in our language, voices concepts corresponding to the property pluralist’s different properties. This gives us an interesting analogue of each version of pluralism outlined above but at the level of concept, rather than property. No such view is an automatic commitment of property pluralism, however, which – as mentioned – is typically combined with conceptual monism, at least in its domain-restricted anti-/realist inflationary moderate form.

But it’s also worth noting, finally, that the entailment does not run the other way either: one can be a conceptual pluralist in any of the many, many ways without holding the analogous property pluralism, for one might be, say, error-theoretic about certain of the concepts. A Kölbelian conceptual pluralist, for example, might hold that the inflationary concept of truth actually fails to pick out any property, and there is only, in fact, the deflationary property corresponding to the deflationary concept. Conceptual pluralism is thus consistent with property monism (or even nihilism). Indeed, *any* combination of concept monism/pluralism with property monism/pluralism is in principle possible. The mind boggles.

## 1.4 Methodology: Analysis and Engineering

### 1.4.1 Role-First Methodology

This thesis is primarily concerned with the property of truth, rather than the concept. A natural view, however, is that the task of concept identification has to precede the task of property identification, since it is the concept that individuates the property of interest. After all, if we give an account of some property that is *not* denoted by the concept, then haven’t we only succeeded in changing the subject?

---

<sup>65</sup> Recently, disjunctivists have resisted this contention (Edwards (2012b: 210-2), Pedersen & Wright (2013b: 103-4)), with Edwards, e.g., arguing that some disjunctions are grounded in genuine similarities among the disjuncts, citing the disjunctive property of *being either crimson or scarlet or vermilion* as an example. To retain a genuine metaphysical similarity here, however, I think commits disjunctivism to the One/Many Gambit (the disjuncts of Edwards’s property are, for instance, determinates of one and the same determinable) and hence to a more robust, inflationary moderate pluralism. A true disjunctivism has no such commitment.

In contrast to this concept-first methodology, I will be pursuing what we might call a “role-first” methodology: identifying explanatory, normative, or expressive roles that we want truth to play, and identifying which property it is that plays that role, and thus which concept *ought* to play that role.<sup>66</sup> I take this to be an instance of what Carnap and Quine called “explication”,<sup>67</sup> what Sally Haslanger has called “ameliorative analysis”,<sup>68</sup> or what has come to be known more recently as “conceptual engineering”.<sup>69</sup>

Conceptual engineers tend to define themselves in opposition to what we might call “conceptual analysts”: the conceptual analyst takes her job to be to identify the nature of the concepts which we in fact associate with certain terms – ‘describ[ing] the actual structure of our thought and talk’, as we might put it.<sup>70</sup> So, for the philosopher of truth, this will be the concept <truth> which is voiced by the predicate ‘is true’. The conceptual engineer, however, worries about the preoccupation with our concepts as we find them. On the one hand, we might be driven by the kind of concerns listed here by Patrick Greenough:

‘Concepts can be defective in manifold ways. They can be intensionally defective: incomplete, confused, unsatisfiable or even incoherent/inconsistent. They can be extensionally defective: too inclusive, too narrow, empty or divided of reference. They can be too complex, too simple, too unspecific or too vague. They can be too parochial or too elitist. They can be redundant or not fit to feature in any explanation. They can be superseded, hackneyed or systematically misapplied. They can be loaded with ideological baggage or serve as ongoing devices for deceit, discrimination, or oppression... [C]onceptual incoherence is just one source of conceptual malfunction...’<sup>71</sup>

Lots of the “malfunctions” Greenough describes probably need spelling out (not least in conjunction with some view of the nature of concepts that allows for them to malfunction like this), but the suggestion is provocative – ordinary language philosophy notwithstanding, philosophers have long complained about the inadequacy of everyday thought and talk when it comes to philosophical and scientific inquiry.<sup>72</sup> But on the other hand, we might just be unimpressed by, as

---

<sup>66</sup> The talk of “playing roles” here may make the position sound like Lynch’s functionalism, but while somewhat similar in spirit, the similarity in approach is superficial. I will not, with Lynch, be assuming that truth is a functional property, nor that there is a single unified functional role for truth, nor that truth has a specifically *functional* role at all, let alone one that can be identified by assembling so-called “platitudes” about it.

<sup>67</sup> See especially Carnap (1950: ch.1) and Quine (1960: §53). For recent discussions, see Maher (2007), Justus (2012), Shepherd & Justus (2015), Brun (2016), and Dutilh Novaes & Reck (2017).

<sup>68</sup> Haslanger (1999, 2000, 2005, 2006).

<sup>69</sup> For some discussion, see Chalmers (2011), Burgess & Plunkett (2013a, 2013b), Burgess (2014), Eklund (2014, 2015, 2017), Plunkett (2016), Greenough (2017), Cappelen (fc). The label is possibly Blackburn’s (2001: 2).

<sup>70</sup> Greenough (2017).

<sup>71</sup> Greenough (2017).

<sup>72</sup> Kevin Scharp (2013) sees philosophy as the study of malfunctioning (in particular, “inconsistent”) concepts, and the proper job of the philosopher as replacing them, if they’re important enough. Scharp himself has

Eklund (2015) puts it, how “navel-gazing” the analyst’s project is. The concepts we happened to have ended up with just might not be that great or interesting, let alone the ones best suited to theorising about complex domains, especially when such theorising serves no direct evolutionary advantage.

The conceptual engineer thus takes her job not to be a *descriptive* one, but a *normative* one: of saying what concepts we *ought* to use, for particular theoretical purposes.<sup>73</sup>

Two points of clarification. Conceptual engineers are sometimes associated with descriptive claims concerning disagreement: Chalmers’s (2011) comments appear in the context of discussing “merely verbal disputes”, while Plunkett & Sundell (2013) explicitly argue that many disagreements are really “metalinguistic negotiations” about how we ought to use words (an outlook which is endorsed by Thomasson (2016); see also Plunkett (2015)). All I want to say here is that such descriptive claims are no part of conceptual engineering as I conceive of it.<sup>74</sup> (No such claim appears anywhere in Carnap, Quine, or Haslanger.) The engineer recommends a philosophical methodology: that we proceed by first identifying a theoretical role – an explanatory, normative, or expressive role, for instance<sup>75</sup> – and then “building” a concept that suits that purpose.

Secondly, Cappelen (fc) thinks of the conceptual engineer as committed to revisionism, to ‘graft[ing] a new meaning upon an old word’, so to speak.<sup>76</sup> But I think it is better to think of the engineer as *tolerating* revision, rather than requiring it. The descriptive question concerning the nature of our actual thought and talk and the normative question concerning the way our thought and talk should be (for certain purposes) are simply two different questions, which may or may not share the same answer. To put the point another way: even if the *concept* we end up with is strictly speaking the same, we may end up with a better or richer *conception* of our subject matter. The engineer should, I think, start from a position of *semantic neutrality* concerning our *actual* thought and talk. How her product relates to our pre-theoretical concepts is then simply a further question.

For instance, in Chapter 5 I ask what property might play the normative role associated with truth of determining which moral judgements we ought to make (in an objective sense), and argue that the expressivist is committed to this being a property I call *weak super-stability*. I thus engineer

---

undertaken this project with regards to truth. Given my end-point, my project here might be profitably viewed as a “replacement” project, akin to Scharp’s. However, I do not start with the idea that <truth> (concept) is inconsistent; rather I start with the idea that truth (property) is *important*, and ask what in the world plays the relevant roles. The spirit of the projects is, I think, very similar, though the letter, and the positive views advocated, quite radically different.

<sup>73</sup> The qualification is important. In the typical case, it will not be central that we start poster campaigns to change how most people use the relevant term. (Though perhaps in some instances. Haslanger, for instance, has *political* purposes – namely, social equality – in mind for her ameliorative analyses of gender and race terms, so perhaps widespread adoption of her analyses is necessary to pursue that end. It is one of Saul’s (2006: 138) criticisms that it’s not obvious when, where, or for whom Haslanger intends her analyses.)

<sup>74</sup> I am unsympathetic with Plunkett & Sundell’s argument, especially with regards to moral disputes. It might be, as Chalmers suggests, that many disputes might be *better off* if we, say, banned certain terms; but this is a million miles from the dubious semantic claim that the arguments are really about something other than what they appear to be about.

<sup>75</sup> The identification of such a parameter I think is vital, though it is often underplayed.

<sup>76</sup> The phrase is from Künne (2003: 26).



the concept <weak super-stability>, which denotes the property which plays the role in question. How this concept and the property it denotes relate to our pre-theoretical concept <truth> and the property it denotes is then a further, complicated, semantic question, on which I need not take a stand.<sup>77</sup>

What of the charge of changing the subject, which can be made against tolerance for revision in meaning, as much as against revision itself? On the one hand, this charge simply seems to beg the question: “the subject” is only being changed if “the subject” is the concept pre-theoretically voiced by the expression ‘is true’, which the engineer ought to reject. On the other, the explanatory, normative, and expressive roles with which we will be concerned here are not being plucked out of thin air, but are the kinds of thing we already expect truth to do. So, insofar as I am interested in finding out what, if anything, in the world plays these roles I am interested in the same questions as philosophers of truth have always been interested in. The view is that the properties identified in this thesis play these roles. Whether or not using ‘truth’ to denote these properties (in the relevant contexts) constitutes changing the meaning of ‘truth’ is a difficult semantic question: if the answer is ‘no’, then there is no objection; but if the answer is ‘yes’ then the view is that we *ought* to change the meaning of ‘truth’ (in the relevant contexts) to denote the properties identified. It is, after all, the explanatory, normative, and expressive roles of ‘truth’ that *matter*. In holding steadfast to my semantic neutrality, I am only refusing to make a substantial and contestable commitment. I do not see how this can be an *objection*.

#### 1.4.2 From Concept Monism to Property Monism?

However, it has been suggested that the existence or otherwise of a unified *concept* of truth might be of *metaphysical* import. Let us finish this chapter by considering this point, which will also allow us to fix some last terminology. Pedersen & Wright (2013b), for example, have argued that if there is a single, platitudinously-identifiable (§1.3.1) concept of truth, then it follows that this concept denotes a single *sparse* property. In particular, even if there is a multiplicity of different truth properties with restricted application, the *disjunctive* property defined by disjoining all these properties (as in *Disjunctive-T<sub>U</sub>* from above, §1.3.2) is a sparse unrestricted property:

‘According to the sparse conception of properties, objects need to be *qualitatively similar* in order to share a property... [But take] the assumptions that the concept of truth is characterized by a collection of [platitudes]... and that T<sub>1</sub>, ..., T<sub>n</sub> qualify as truth properties in virtue of satisfying these [platitudes]. For if the assumption holds, then satisfying the

---

<sup>77</sup> The rough view to which I am sympathetic is that our use of ‘truth’ “tracks” this property in the moral domain, but that the concept <truth> itself is inconsistent. The semantics of inconsistent concepts is a very complicated issue I cannot touch on here. See especially Scharp (2013: ch.9), Eklund (2002a; 2017), and the discussion at §8.1.2.

[platitudes] is sufficient to satisfy this constraint of qualitative similarity that the generic disjunctive property has to satisfy in order to qualify as metaphysically legitimate [i.e., sparse].<sup>78</sup>

The thought seems to be this. Take a “platitude” like Lynch’s “Norm of Belief”: it is *prima facie* correct to believe that p if, and only if, the proposition that p is true. Suppose, then, that it is *prima facie* correct to believe those propositions in S<sub>1</sub> iff they correspond, and *prima facie* correct to believe those propositions in S<sub>2</sub> iff they are superwarranted. Then correspondence and superwarrant have a “qualitative similarity”: they are properties that make (distinct sets of) propositions *prima facie* correct to believe. So, the property of either (being in S<sub>1</sub> and) corresponding or (being in S<sub>2</sub> and) being superwarranted is a sparse unrestricted property: namely, the property that makes propositions *prima facie* correct to believe. (In a similar spirit, Lynch (2004: 390) claims that ‘truth is *general* cognitive norm.’<sup>79</sup>)

Now, I will have cause to pause on the normativity of truth for belief later (§4.3). But on its own terms, the argument is suspicious. Rather than being an argument *for* there being an unrestricted sparse truth property, it seems to simply *assume* that there must be a single sparse property that norms all belief. (Analogous points can be made for the other platitudes.) This is to beg precisely the point at issue. For the restricted pluralist may want to deny precisely this point, and maintain that some beliefs are normed by one property, others by another. Indeed, that seems to be *exactly* the pluralist’s claim, if she grants that these properties satisfy the platitudes for different sets of truthbearers. To make the argument stick, Wright & Pedersen should give us some reason to think that all beliefs are normed by the same property. (E.g., when introducing the platitude, Lynch (2009: 11) stresses the idea that the norm of truth is *constitutive* of the nature of belief, meaning we might end up with a bifurcation in our metaphysics of belief unless there is a unified norming property. See §4.3.2.) But that is a *further* argument, and does not follow merely from the “platitude”.<sup>80</sup>

To guard against a potential source of confusion: property pluralists often introduce the view by saying that there is more than one way of *being true*, which might imply that there is a single property – *being true* – which there is more than one way of being.<sup>81</sup> Similarly, I will often say that truth “consists in” different properties for different truthbearers (indeed, I have already been talking

---

<sup>78</sup> (2013b: 103-4). See also Edwards (2012b: 211-2).

<sup>79</sup> Lynch is raising this as a problem for the view that ‘truth’ is ambiguous, i.e., conceptual pluralism. It is striking that, in his (2009: 62-7), he does not explicitly raise it against Wright’s property pluralism as he does all the other problems he raises for conceptual pluralism (although see (2009: 65)).

<sup>80</sup> See also Eklund (2015), who raises the idea that truth is the aim of belief to suggest that <truth> may be a “conceptual fixed point”; i.e., incapable of being “re-engineered”. Eklund’s primary target, however, is not someone who says that truth consists in different things for different truthbearers, but those who would replace <truth> with some other concept. Note that Wright himself is sceptical of the argument quoted above; see Wright (2012: 95) for his reasons.

<sup>81</sup> Some philosophers have tried to leverage this into an argument for monism, e.g., Tappolet (2000: 384), Lynch (2005: 42; 2009: 58-9).

in this way), which might imply that there is a single property – *truth* – which consists in different things. (I will sometimes speak of different truthbearers being “apt for” different truth properties.)

To cancel this implication, let me stipulate the following terminology. If  $T_i$  is *the truth property relevant for* a truthbearer ‘p’, then for any predicate ‘ $T_i$ ’ that denotes  $T_i$ , ( $T_i$ -In) and ( $T_i$ -Out) are valid, where the single-quotes here serve as an operator that takes sentences to the relevant primary truthbearer:

( $T_i$ -In)	p. _____ ‘p’ is $T_i$ .	( $T_i$ -Out)	‘p’ is $T_i$ . p.
--------------	----------------------------	---------------	----------------------

To be a *truth property* is to be the truth property relevant for some proposition, in this sense.<sup>82</sup> A proposition’s truth *consists in* a truth property, and it is *apt for* that property, just in case that property is relevant for the proposition, in this sense. All truth properties, then, have this “in common”. One might thus think that we can define a univocal truth predicate via generic introduction and elimination rules of this kind. I think this is probably right *in a sense*, and I am even tempted by the idea that the concept voiced by our predicate ‘is true’ is just such a concept (a kind of inferentialism about <truth>), though I am also inclined to think that the concept so defined is inconsistent (see Chapter 8). The point of the above argument, however, is that *even if* there is such an unrestricted *concept*, we cannot assume that there is a single sparse property that is relevant for all truthbearers without further argument.

## Conclusion

In this opening chapter, I have surveyed a variety of views on truth with three core dialectical purposes in mind: (i) rendering salient the views to which I am responding, (ii) introducing core elements of my own view, and (iii) situating my view within the literature. I have also presented a sophisticated taxonomy of truth pluralism and explained my methodology. In the next chapter, I critique extant motivations for anti-/realist pluralism, before developing my own such argument in Part II.

---

<sup>82</sup> Note that I am only stipulating terminology here. I am making no profound claim about *what it is* to be a truth property.

# CHAPTER 2

## SCOPING OUT TRUTH PLURALISM

---

This chapter concerns motivations for domain-restricted anti-/realist pluralism about the nature of truth, and in particular the “scope problem”. In one way or another, the scope problem tries to draw on the diversity of truths – and, in particular, localised realism and anti-realism – to motivate pluralism about the nature of truth. But the argument is all-too-rarely left impressionistic and intuitive. In this chapter, I critically evaluate the “ontologically-driven” versions of the scope problem that have dominated the literature hitherto. While I proceed by offering detailed critical analysis of particular incarnations, my goal is by the end to have highlighted a structural, methodological inadequacy in this approach, and thus set the stage for my own “metasemantically-driven” and direct argument for pluralism in Part II.

### 2.1 Proto-Scope Problems

#### 2.1.1 *Proto-Scope Problems*

In its most intuitive – or, perhaps, naïve – form, the scope problem involves gesturing at all the truths and saying “What on Earth could be in common among all of *them*?” It is the philosophical equivalent of taking a biologist to a zoo, pointing at all the lifeforms contained therein – the sloths, the hagfish, the palm trees, the turtles, the cockatoos, the spiders, the dolphins – and in wide-eyed astonishment at the sheer diversity, daring them to say what *life* consists in. We can, perhaps somewhat uncharitably, see this in the “problem of the common denominator” that Gila Sher extracts from Kant:

‘In plain words and using contemporary terminology, we can explain the prima-facie valid core of Kant’s argument as follows: Consider two truths of the same syntactic structure but altogether different subject matters, say, ‘ $2 < 10$ ’, and ‘John loves Mary’ (assuming it to be true). In spite of their syntactic similarity, the substantive truth conditions of these two sentences appear to have nothing in common. The truth of ‘ $2 < 10$ ’ is a matter of such things as the positions of 2 and 10 in the natural number series, the non-existence of

surjective (“onto”) mappings from sets with 2 elements to sets with 10 elements, and so forth; the truth of ‘John loves Mary’, in contrast, is a matter of John’s feelings towards Mary, John’s behavior regarding Mary, John’s brain states involving a representation of Mary, and so on. The truth conditions of these two sentences are altogether different, and an informative theory of truth must account for these differences. But a general theory of truth must abstract from these differences. Ergo: a contradiction. *A substantive theory of truth is not general; a general theory of truth is not substantive.*’ (Sher 1998: 139-40)<sup>1</sup>

This intuitive argument does not pass much rigorous muster, especially once we consider the actual inflationary theories that have been put on the table. For example, on the face of it, the correspondence theory sidesteps the supposed generality/specificity trade-off by maintaining that different truthbearers stand in *the same relation* (substantive generality) to *different truthmakers* (informative specificity).<sup>2</sup> In this form, the scope problem does not so much present a *problem* for the monist as it does emphasise the scope of her project. This might be intimidating, but philosophical inquiry is often intimidating, especially at the outset.

The scope problem more usually appears in a slightly more fleshed out – but still impressionistic – form, exemplified here:

‘...some theories – such as correspondence theories – seem intuitively plausible when applied to truths about ladders, ladles, and other ordinary objects. However, those theories seem much less convincing when applied to truths about comedy, fashion, ethical mores, numbers, jurisprudential dictates, etc. Conversely, theories that seem intuitively plausible when applied to legal, comic, or mathematical truths – such as those suggesting that the nature of truth is coherence – seem less convincing when applied to truths about the empirical world.’ (Pedersen & Wright 2016: §2)<sup>3</sup>

This idea certainly has some intuitive pull, but, so understood, it is no wonder the scope problem is rarely developed in detail: it requires arguing for (i) the localised “plausibility” of truth

---

<sup>1</sup> The scope problem can also be found in this form in Lynch (2004: 385; 2005: 32-3), Sher (2004: 8-10), and Wright (2005: 2).

<sup>2</sup> At times, Sher (or, perhaps, Kant or Sher’s Kant) seems to think that, to give an account of the nature of truth for a truthbearer <p>, the correspondence theorist needs to include the relata (i.e., <p> and p) that *stand in* the relation (on that occasion, as it were). This would be a straightforward confusion. One might worry what relation could tolerate such diversity in relata (this is, in essence, where more developed versions of the problem aim to strike), but (i) it is contentious that we require any such diversity (the worldly relatum may always be a *fact*, albeit one composed of radically different things in different cases); and (ii) one needs to *show* that no such relation can.

<sup>3</sup> Other presentations that go no further than stating the domain-variable im/plausibility without even gesturing at a defence or explanation include Lynch (2006: 63; 2013b: 22) and Wright (2010: 270-1; 2012: 92). Note Wright’s (2012: 93) scepticism: ‘pluralists (as a group) have done amazingly little to expound on the scope problem or to go far beyond the usual glosses on it’. I agree.

property  $T_1$  for set of truthbearers  $D_1$ , (ii) the localised “implausibility” of  $T_1$  for the disjoint set of truthbearers  $D_2$ , (iii) the localised plausibility of distinct truth property  $T_2$  for  $D_2$ , and (iv) the localised implausibility of  $T_2$  for  $D_1$ . This is an enormous task. Note that neither of the positive theses (i) and (iii) entail, nor even necessarily lend support to, the negative theses (iv) and (ii): showing that one theory of truth is plausible for one set of truthbearers does not automatically entail that no other theory of truth is plausible for said truthbearers. And “plausibility” is something of a weasel-word here. For, as pluralists themselves sometimes admit, it is not as though the *only* problems that traditional theories of truth have faced are problems of scope. Given that any theory will face criticisms on its own terms (call them “theory-internal” as opposed to “domain-specific” problems), we really need to be told why the domain-specific problems that  $T_1$  faces with respect to  $D_2$  renders  $T_1$  “implausible” for that domain, but the theory-internal problems do not render  $T_1$  “implausible” wholesale, including in  $D_1$ . That is, the pluralist ought to argue that the domain-specific problems are fatal to  $T_1$ ’s prospects (in  $D_2$ ) in a way that the theory-internal problems are not.<sup>4</sup> As it is, the scope problem is most usually presented only in passing, the arguments for (i) and (iii) and the domain-specific problems in favour of (ii) and (iv) only gestured at.

Even granting (i)-(iv), a plethora of non-pluralist options are available. The monist might dig her heels in, and deny that those truthbearers for which her preferred theory is implausible can be true. The realist or correspondence theorist in particular may use the implausibility of her theory in a domain as a motivation for error theory. Or, she might allow that the truthbearers are true, but endorse a hermeneutical fictionalism in the relevant domain (so, the relevant truthbearers will only fail to be true when (if) they are (can be) stripped of their pretence). Or, she might deny that they are truth-apt.<sup>5</sup> Lynch (2009: 35) suggests that the existence of such options only shows the force of the scope problem; but the flipside of this is that truth pluralism is born into a world of fierce theoretical competition. And there are non-inflationary-monist options too. The global deflationist (§1.2.1) can allow that all the relevant truthbearers can be true, but deny that there is anything their truth consists in. Pluralists often press anti-deflationary considerations, but unless she shows deflationism to be in some way incoherent, in this context she needs to show not just the global inadequacy of deflationism, but also its local inadequacy for *each* of the relevant domains about which she is an inflationist. For the partial inflationist – who maintains that all truthbearers can be true in a deflationary sense, but only some can be true in an inflationary sense (§1.3.2) – may allow that  $T_1$  is right for  $D_1$  without allowing that a rival inflationary theory is right for  $D_2$ . And, of course, the anti-/realist pluralist needs to give us some reason to prefer her view to correspondence pluralism: the view that different truthbearers correspond in different ways (§1.3.2). Finally, basing pluralism

---

<sup>4</sup> Wright (2012: 102) worries that ‘pluralists seem to be masochistically shouldering everyone else’s burdens – in effect, saddling themselves with each objection or problem which harangues the various candidate ways of being true that pluralists posit.’

<sup>5</sup> Pluralists sometimes push general worries about these strategies, but given that each of the pluralist’s truth properties, and pluralism itself, will face their own theory-internal problems, this seems dialectically unsustainable.

on (i)-(iv) leaves it vulnerable to the possibility that there is some omni-plausible truth property that we simply haven't identified yet. In the interests of charity, I will set these rival options aside for the most part in what follows. (Note that my own preferred version of the "scope problem", developed in Part II, does not go via this localised im/plausibility strategy.)

### 2.1.2 Explaining Anti-/Realist Intuitions

Wright (1998b: 58) suggests a different motivation, though one similar in spirit:

'A pluralistic conception of truth is also philosophically attractive insofar as an account which allows us to think of truth as constituted differently in different areas of thought might contribute to a sharp explanation of the differential appeal of realist and anti-realist intuitions about them. But I acknowledge, of course, that more detail and a sharper theoretical setting is required for the proposal before it can really be clear that it makes genuine sense, let alone possesses merit.'

To avoid confusion, this motivation should be distinguished from the closely related idea that truth pluralism is 'how we should best understand the contrast between so-called realist and anti-realist views concerning different areas of thought and discourse, and [is] how the debate between them might most effectively be persecuted.' (Wright 1992: 1)<sup>6</sup> The latter thought would concern something like philosophical methodology or interpretation: that realists and anti-realists ought to be disputing about the nature of truth in the relevant domain, or that we should *understand* realists and anti-realists as (fundamentally?) disputing the nature of truth in different areas of thought, and I'm not sure either was ever one of Wright's views. Rather, Wright sees a number of different cruxes as being productive fodder for anti-/realist disputes: cognitive command, width of cosmological role, the Euthyphro contrast, evidence transcendence. In different domains, Wright suggests, our intuitions may come down on different sides of this cruxes. The idea that the nature of truth is different for these domains may, then, be *part* of the best explanation of our domain-variable intuitions.

As Wright partly concedes in the second sentence of the above quote, this leaves us with an awful lot of work to do. We first have to gather the data regarding intuitions that is to be explained (a project to which Wright himself has, of course, contributed considerably) and, naturally, make the case that these intuitions are to be trusted. We then have to work out the pluralistic view, or the explanation of which the pluralistic view is a part. We then have to work out its competitor explanations. And after all that we have to weigh them against each other. This is an enormous

---

<sup>6</sup> Pettit (1996: 885) seems to understand Wright this way.

project, and not one we can hope to undertake here.<sup>7</sup> However, it is clear that any such motivation for truth pluralism will have to face what has been known as the “Quine-Sainsbury objection”:

‘...even if it is one thing for “This tree is an oak” to be true, another thing for “Burning lives cats is cruel” to be true, and yet another for “Buster Keaton is funnier than Charlie Chaplin” to be true... we get a better explanation of the differences by alluding to the differences between trees, cruelty and humour. We could thus be minimalist about what it is for sentences about comedy to be true, some species of realist (e.g. one committed to Cognitive Command) about what it is for sentences about morals to be true, and a stronger species of realist (adding, perhaps, Wide Cosmological Role) about what it is for sentences about familiar material objects to be true... The dispute between anti-realists and realists need not bear essentially on the nature of truth, but instead on the nature of the subject matter introduced by the sentences to which the predicate “true” is applied.’ (Sainsbury 1996: 900)<sup>8</sup>

Perhaps we can think of ‘truth’ as voicing a unified concept or denoting a unified property, and explain domain-variable anti-/realist intuitions by appeal to the differences in their *subject matter*, i.e., differences in what the truthbearers in the relevant domain are *about*, rather than the nature of truth for the relevant truthbearers. Note that we need not appeal to a deflationary conception of truth to make this move (though those – like Blackburn (2013a) and Dodd (2013) – who press this argument typically have this in mind). The correspondence theorist, for instance, can appeal to variability in the nature of the worldly relata of the correspondence relation; the “truthmakers”.<sup>9</sup> Also note that all the other theoretical options outlined above will compete here.

---

<sup>7</sup> A related motivation is that truth pluralism might explain why there has been, historically, a lack of consensus as to the nature of truth, and why ‘the traditional debate about truth turned out to be sterile and incomplete.’ (Wright 2013: 123; this is also prominent in Lynch 2001: 723). Perhaps. Again, to lean heavily on this motivation, the pluralist really needs to spell out the proposed explanation in detail, and make the case that it is the best explanation. Compelling evidence might involve showing that the historical champions of different theories really have placed an over-emphasis on a particular range of truths in their discussions; in particular, those in the pluralist’s preferred domains (coherence theorists on mathematics, say). I am not optimistic that such a prediction would be borne out.

<sup>8</sup> See also Quine (1960: 118-9). Both Blackburn (2013a) and especially Dodd (2013) have recently argued in this vein that the pluralist is guilty of a kind of “double-counting”: that we can acknowledge and detail the differences between domains without further appeal to differences in the nature of truth.

<sup>9</sup> Asay (2016) has recently pressed this idea. Asay also puts forward an argument against truth pluralism, and it’s worth pointing out (since this is a recent publication in a prominent journal by an excellent philosopher) that this argument rests on a simple equivocation. Asay interprets truth pluralism as a view about *truthmaking* (2016: 3): that different truthbearers are *made true* by different properties. He then argues, as truthmaker theorists are wont to do, that what makes a truthbearer like <Socrates is a philosopher> true is not a property the truthbearer possesses, but a state of affairs (a fact-like thing): Socrates exemplifying the property of being a philosopher. Setting aside other objections, there are two notions of “truthmaking” in play here, which have long been distinguished. (I haven’t traced back how far, but it is made in Marian David’s 2005 review of Vision (2004) – see <http://ndpr.nd.edu/news/veritas-the-correspondence-theory-and-its-critics>; see David (2016: §2.2) for a particularly clear presentation.) To say that a truthbearer is *made true* by a property might mean that its truth *consists in* that property: truth pluralists are pluralists about this. There is nothing peculiar about this



Let me not overstate the case I am making. Domain-variable anti-/realist intuitions, domain-specific problems for particular theories, the sterility of and lack of consensus in the historical debate, and the sheer diversity of truths give us good reason to be *interested in* truth pluralism, as a theoretical possibility. Perhaps, all things considered, this will be the best option. But this kind of philosophical motivation is cheap: for the open-minded inquirer, logical space mongering is practically always a virtue. Given the above observation that the theoretical competition here is fierce, what we're after is something stronger: a reason to think that truth pluralism is *right*.

But note that even the advocate of the Quine-Sainsbury objection must allow that the pluralist *might* be onto something here: the very differences that she herself appeals to may “make for” differences in the nature of truth between domains. What we need, however, is an *argument* that this is so. Such an argument will proceed in two steps: (1) saying what the relevant difference between the domains is, and (2) saying why this difference motivates truth pluralism. Not just any answer to (1) will provide a solid ground for (2). For a silly example, consider an orthographic difference: the fact, say, that in written English we represent numbers using numerals sometimes composed of other numerals ('1', '53151', '6.419') while we use words composed of letters from the Roman alphabet ('table', 'coffee mug', 'World War One') to represent tables, coffee mugs, and wars.<sup>10</sup> This difference gives us no reason to think the nature of truth varies between domains. Pluralists are generally impressed by differences between domains, but it is incumbent on the pluralist to say *what* differences between domains make for differences in the nature of truth. In doing so, she will not be, and need not be, neutral on substantive questions concerning similarities and differences between different domains. It is not the job of the pluralist qua pluralist to motivate an answer to (1), which can be done independently. Her job is to answer (2): to give a convincing case for why the differences assumed in answer to (1) motivate truth pluralism.

We will see that pluralists have typically had a particular kind of answer to (1) in mind, which has been nicely captured by Pedersen:

‘...what determines features of the truth property relevant to a given proposition is the nature of its subject-matter, i.e. the part of reality that it is concerned with.’ (2014: 275)

This may sound neutral, but it is not: on Pedersen's view, what motivates truth pluralism are *ontological differences*: differences in kind between the objects and properties – the “parts of reality” –

---

relation to truth: something is *made water* by having the molecular structure of H<sub>2</sub>O. In the sense in which Asay requires truthmaking, however, it is a relation peculiar to truth: it is an asymmetric relation that a state of affairs bears to a truthbearer, perhaps necessitating its truth. It is simply consistent with the correspondence theory that truths have truthmakers in this sense (indeed, it is sometimes taken to be *constitutive of* the correspondence theory – see e.g., Dodd (2000: ch.1) – though I agree with Asay (2013) that this is a mistake). And insofar as an anti-realist theory of truth for ‘a is F’ can admit that a exemplifies F-ness, even these theories can acknowledge a “truthmaker”. I intend to develop this response in detail elsewhere.

<sup>10</sup> This difference is exaggerated of course, since we often do use words composed of letters to represent numbers ('one', 'fifty-three thousand, one hundred, and fifty-one', 'six point four one nine').

that are the subject of discourse in different domains. By contrast, my own preferred view utilises differences at the level of *thought and language*.<sup>11</sup>

## 2.2 Ontologically-Driven Scope Problems

### 2.2.1 Cotnoir & Edwards: Truth, Being, and Grounding

In a recent paper, Aaron Cotnoir and Douglas Edwards have offered an intriguing motivation for truth pluralism which assumes an underlying *ontological* pluralism.<sup>12</sup> Ontological pluralism, as they are interested in it, assumes an inflationary answer to the question of what it is for something to exist, but argues that the nature of existence varies for different kinds of object. For some objects:

(Alexander’s Dictum)                      To exist is to have causal powers (2015: 119)

while for other objects:

(Neo-Fregean Principle)                      To be is to be the referent of a singular term that appears in a true sentence. (2015: 120)

For instance, “concrete” objects like tables, the Eiffel Tower, and tennis balls might exist in the former way, while abstract objects like numbers might exist in the latter. Following Cotnoir & Edwards, let’s say that objects that exist in virtue of having causal powers have BEING<sub>1</sub>, while objects that exist in virtue of being the referents of singular terms that appear in true sentences have BEING<sub>2</sub>.

Given this assumption, Cotnoir & Edwards contend, ‘one has a natural route to truth pluralism’ (2015: 128). The argument is as follows. Call the property that true sentences have TRUTH.<sup>13</sup> Suppose that BEING<sub>1</sub> grounds the truth of some sentences; so BEING<sub>1</sub> grounds TRUTH.<sup>14</sup> And suppose further that the truth of some sentences grounds BEING<sub>2</sub>; so TRUTH<sub>i</sub> grounds BEING<sub>2</sub>. This latter claim just seems to be part of the Neo-Fregean Principle, as an account of existence. Now, if TRUTH<sub>i</sub> = TRUTH<sub>j</sub>, as monists about truth contend, then by the transitivity of grounding, BEING<sub>1</sub> grounds BEING<sub>2</sub>. But if BEING<sub>1</sub> grounds BEING<sub>2</sub>, then BEING<sub>1</sub> is more

---

<sup>11</sup> Ball’s (2017) proposal is superficially similar to my own on this front, but he works within a very different framework to that developed here. Discussion would unfortunately take us too far afield.

<sup>12</sup> Cotnoir & Edwards (2015: 128-30).

<sup>13</sup> We’ll talk about sentences here to be consistent with Cotnoir & Edwards’s discussion.

<sup>14</sup> Cotnoir & Edwards seem to see this as part of the claim that some entities have BEING<sub>1</sub>, though I am not sure why. This ontological claim by itself says nothing about truth. But it is a plausible enough claim within their framework, so I won’t take issue with this premise here.

fundamental than BEING<sub>2</sub>, and this violates the equi-fundamentality of BEING<sub>1</sub> and BEING<sub>2</sub> that is constitutive of ontological pluralism. So, by reductio, TRUTH<sub>i</sub> must be distinct from TRUTH<sub>j</sub>; that is, the property possessed by those true sentences that is grounded in BEING<sub>1</sub> is distinct from the property possessed by those true sentences that ground BEING<sub>2</sub>.

Within Cotnoir & Edwards’s framework, I don’t think this argument works. The equi-fundamentality claim I find odd: the first (and only) time it appears in the paper is in this argument, and it seems not just unsupported by, but inconsistent with the ontological pluralism presented. Consider the analysis of BEING<sub>2</sub>: the Neo-Fregean Principle (hereafter NFP). This explains the existence of certain entities, perhaps numbers, in terms of the truth of certain *sentences*. For these sentences to be true, *they* must exist. What kind of existence do these sentences have? If they have BEING<sub>1</sub>, then BEING<sub>2</sub> is grounded in BEING<sub>1</sub>, contra the equi-fundamentality claim. But if sentences *themselves* have BEING<sub>2</sub>, then we are off on a vicious regress. The natural way to stop the regress is to allow that some sentence has BEING<sub>1</sub>, and then we are back on the previous horn. Cotnoir & Edwards, at this moment at least, seem to think that the equi-fundamentality claim is necessary for their variety of ontological pluralism (2015: 129), in which case this dilemma refutes the view. But, for what it’s worth, it seems to me no part of the general ontological pluralist picture they present that BEING<sub>2</sub> should not be grounded in BEING<sub>1</sub> in the way just suggested; what is important is that some objects exist in one way, and that other objects exist in another (and that, say, existing in one way is not merely a *way* of existing in the other). Either way, the argument from ontological pluralism to truth pluralism fails.

Can we get a similar argument going in the same spirit? What is critical for the argument is the claim that the truth of a sentence sometimes grounds, and is sometimes grounded in, the existence of the referent of its singular term(s). If one can argue for *that*, perhaps on the basis of ontological pluralism, then it is a short step to truth pluralism.<sup>15</sup>

Suppose, then, that one is antecedently committed to the correspondence theory of truth for some sentences. This might be because one is sympathetic with the broadly realist picture that there are objects with BEING<sub>1</sub>, and that the truth of sentences about these objects is grounded in the existence of these objects, in the way suggested by the correspondence theory. Presupposing this much, of course, makes the argument much weaker. But imagine we have an independent argument for the correspondence theory for some sentences (perhaps like that provided in Chapters 3-4).

Now we ought to note that NFP is *incompatible* with the correspondence theory of truth. To see this, imagine we rewrite the principle thus:

(NFP-C)            To be is to be the referent of a singular term that appears in a sentence that corresponds.

---

<sup>15</sup> Cotnoir & Edwards (2015: 129-30) consider some responses, but we’ll set these aside.

The problem, of course, is that *correspondence* is to be understood in part in terms of the object that is the referent of the sentence's singular term (§1.2.2), making NFP-C, as an account of the nature of existence, circular. So, one might think, if one can be independently convinced of NFP, one is compelled to reject the correspondence theory for the relevant sentences, and thus compelled to be a truth pluralist.

The core problem with this way of motivating truth pluralism, as I see it, is the following. The truth pluralist will think that truth for the sentences that ground the existence of entities with BEING<sub>2</sub> is some property other than correspondence, like, say, superwarrant. Given this, it is hard to see how one might possibly motivate NFP, which appeals to the *truth* of the sentences, over the related principle:

(NFP\*) To be is to be the referent of a singular term that appears in a superwarranted sentence.

But NFP is critical for the argument. The global correspondence theorist can admit NFP\*: the number 5, she might think, exists because 'the number 5 is prime' is superwarranted; but 'the number 5 is prime' is *true* because the object referred to by 'the number 5' (i.e., the number 5) instantiates the property ascribed by 'is prime'. The key point here is that the ontological principle that Cotnoir & Edwards's argument appeals to does not just ground the existence of the relevant objects in the sentence's having *some* property, but in its *truth*. On the one hand, one may worry that appealing to this principle hereby *presupposes* truth pluralism, insofar as this principle is incompatible with the correspondence theory, which we are assuming.<sup>16</sup> On the other, my present challenge is to ask how, by the pluralist's own lights, NFP can be motivated over the rival principle NFP\*. Without this, the argument does not go through.

One possible response is to point out that other sentences, like those concerning tennis balls, can be superwarranted without the relevant objects existing; or connectedly that *truth* requires the existence of the relevant objects while superwarrant does not. But "requires" in what sense? It is no part of the *nature* of truth that the singular term has a referent *if truth is superwarrant*. It is only on a correspondence theory that this is so. Furthermore, both sides agree that superwarrant is sometimes sufficient for existence, sometimes not. (This is an oddity of the view.) While the advocate of NFP will maintain that it is only when superwarrant *is* truth that it is sufficient for existence, the advocate of NFP\* will maintain that this gets the order of explanation the wrong way around. It is because superwarrant is only sufficient for existence in some cases that in *those* cases

---

<sup>16</sup> Of course, it does not by itself presuppose pluralism, but only the falsity of the correspondence theory for some sentences; but, as just suggested, I do not see any way of getting the argument going *without* presupposing the correspondence theory for some sentences. In conjunction with this assumption, the principle presupposes pluralism.

the sentences are true. Similarly if the advocate of NFP tries to draw on the claim that truth is “conceptually” or “analytically” sufficient for existence, where superwarrant is not. The correspondence theorist can simply agree, since on her view reference (and hence existence) is necessary for truth; she just puts the order of explanation the other way around, consistently with NFP\*. One might just think the combination of NFP\* and the correspondence theory is weird, of course; but the view only differs from the alternative in refusing to bifurcate the nature of truth as well as existence, so it can hardly be weirder than its competitor.

I do not wish to overstate the case: Cotnoir & Edwards have shown that the kind of ontological pluralism they are interested in and truth pluralism make for natural bedfellows, which is their primary goal. However, I disagree that there is any straightforward *argument* from the former to latter. In particular, my core worries about their argument are that it smuggles in truth pluralism by the back door by assuming an account of existence that is grounded in truth; and that by the truth pluralist’s own lights it is difficult to see what could motivate this principle over an alethically innocent one.

### 2.2.2 Pedersen: Mind-In/dependence and Metaphysical In/completeness

A similar view has also been advocated by Nikolaj Pedersen (2014). Pedersen, like Cotnoir & Edwards, is primarily concerned to show that alethic pluralism “fits well” with a kind of ontological diversity (in Pedersen’s mouth, “metaphysical” pluralism). Despite suggesting that ‘the basic motivation for the pluralist package is metaphysical’, he (somewhat disappointingly) does not offer any direct argument from ontological to truth pluralism, saying that ‘[t]he present paper is not the place to defend or argue in favour of the fundamental role attributed to metaphysical pluralism.’<sup>17</sup> (2014: 275)

What Pedersen does suggest is that his ontological pluralism can *explain* the “metaphysical completeness” of some domains and “incompleteness” of others, which he argues is distinctive of truth pluralism. Ontological pluralism, for Pedersen, is the idea that some objects exist *mind-independently*, while other objects exist *mind-dependently*, in the sense that they would cease to exist if all minds ceased to exist.<sup>18</sup> (Note that Pedersen’s formulation explicitly concerns objects, rather than properties; he thus seems to have the mathematical case at the forefront of his considerations. Note also that his pluralism need not be as radical as Cotnoir & Edwards’s: it seems consistent with a deflationary attitude towards the nature of existence itself.) A domain is “metaphysically complete” for Pedersen just in case any state of affairs in that domain either obtains or does not obtain. A metaphysically incomplete domain is not metaphysically complete.

---

<sup>17</sup> By the ‘pluralist package’, Pedersen means not only ontological and truth pluralism, but also logical pluralism.

<sup>18</sup> Cotnoir & Edwards (2015: 118), Edwards (2011: 29), Lynch (2001: 724; 2004: 385), Pedersen (2006: 102; 2010: 93), and Wright & Pedersen (2010: 210) all suggest that the mind-in/dependence contrast might be central to the scope problem. Pedersen’s discussion is one of the only real attempts to say *why*.

Pedersen claims that

‘the following idea has some pull:... For any mind-independent object *a* and mind-independent property *F* in brute physical reality, it is determinate whether *a* is *F* or it is not. What this idea amounts to is that brute physical reality deals in sharp boundaries.’ (2014: 273-4)

This strikes me as much more controversial than Pedersen lets on, given, for instance, the existence of vagueness. Ajax the cat is, let’s assume, mind-independent, as is the property of baldness. But it may not be the case that it is *determinate* whether or not Ajax is bald, because he may be a borderline case. At best, Pedersen seems to be ruling out the possibility of ontic vagueness by fiat, and at worst he is presupposing epistemicism.<sup>19</sup> Of course, Pedersen is only suggesting that the idea “has some pull”; but insofar as the picture he is pushing relies on this principle, this kind of worry cannot be ignored. One might suggest that Pedersen does not have properties like *baldness* in mind when he suggests that “brute physical reality” is determinate: while our language or thought uses vague terms or concepts, the properties that appear in brute physical reality are not the ones ascribed by such predicates. But given that vagueness infects the vast majority of our language and thought, including all those parts that are *prima facie* concerned with mind-independent properties, it will turn out that Pedersen is not concerned with the vast majority of properties we might have thought – and which he explicitly states – he is concerned with: ‘moons and mountains, rocks and rivers’ (2014: 273) will be no part of it. (And what to say about those truthbearers concerning mind-independent properties that *do* feature vague terms or concepts?)

More convincing is the move from mind-dependent existence to metaphysical incompleteness. Pedersen considers the mathematical intuitionist, for whom the existence of mathematical objects (with certain properties) is a matter of being “constructed” via a demonstration or proof. Given a certain mathematical hypothesis, like the Twin Prime Conjecture (that there are infinitely many pairs of primes two apart), there may be at some moment neither a proof of it nor its negation – this is, in fact, so. Given this, the intuitionist can maintain that there is no fact of the matter, at present, as to the status of the Twin Prime Conjecture.<sup>20</sup>

It’s not obvious, then, that domain-variability in mind in/dependent existence of the relevant objects offers a good explanation of domain-variability in metaphysical in/completeness, or even that the latter is an explanandum. *Even granting*, however, this explanation, the difficulty will be in moving from this to pluralism about truth. Pedersen takes such metaphysical in/completeness to be ‘characteristic of respectively realist and anti-realist truth.’ (2014: 274) Even this is not obvious.

---

<sup>19</sup> For ontic vagueness, see Williams (2008); for epistemicism, see Williamson (1994b).

<sup>20</sup> I am sliding here between metaphysical talk of the states of affairs of the relevant domain and semantic talk of the truth of the conjecture about the domain, but we need only grasp the gist of the argument here. Pedersen (2014: 274) provides more details, though I think the argument could use more fleshing out.

Superwarrant is *compatible* with incompleteness, since neither a proposition nor its negation may be warranted in a state of information; but the relevant domain may, nonetheless, be metaphysically complete, and correspondingly every proposition either true or false. More troubling is that Pedersen assumes metaphysical completeness is a feature of the correspondence theory of truth. Given the possibility of vagueness, I see no reason why this should be so. Indeed, it's clear that Pedersen's line of thinking here does not so much concern the truth properties of *atomic* propositions in the relevant domains, but the truth properties of *negations*. Where he assumes that the relevant truth property for a negation in a domain where superwarrant is relevant for atomics is still superwarrant (2014: 264), he assumes that the relevant truth property for a negation in a domain where correspondence is relevant for atomics is *the absence of truth (correspondence) of the negand* (2014: 267). This assumption might be plausible, but it needs to be argued for; it is this assumption that does all the heavy lifting!

Ultimately, however, Pedersen will also face an analogous problem to the one presented against Cotnoir & Edwards above. Even if we think that the *existence* of the relevant objects and properties in a domain is down to the relevant propositions being proven, we might still think that the *truth* of said propositions consists in its corresponding with the objects that so exist. That is, we might put the explanatory order from proof to existence to truth, rather than presupposing that proof is truth, and thus putting the order from proof/truth to existence. As argued above, it is difficult to see, by the pluralist's own lights, what might support the latter over the former.

\*\*\*

There is general point to be made here. I think a large part of the intuitive pull of pluralism comes from starting with the correspondence theory for some domains, and then feeling like the objects and properties of other domains may be *absent* (because they're weird, or we can't point to them, or whatever), which means they cannot be relata of the correspondence relation.<sup>21</sup> Nonetheless, we want the propositions of, say, mathematics to be true! Hence pluralism. The problem is that, if ' $2 + 3 = 5$ ' is true, then (at least on the face of it), the objects it talks about *exist*. It would be very odd to allow that ' $2 + 3 = 5$ ' is true and yet all the singular terms involved fail to refer to anything. And if they exist, then they exist to be corresponded with; there is no absence of relata in the relevant domain.

Cotnoir & Edwards and Pedersen try to make good on this intuition without falling into this trap by claiming that the objects exist, but have a different *kind* of existence. The problem is that, if the objects exist *at all*, then they exist to be corresponded with. This is why Cotnoir & Edwards

---

<sup>21</sup> In one of his "in passing" presentations of the scope problem, Lynch (2008: 122) explicitly endorses this line of thinking: 'We judge radically distinct kinds of propositions to be true – ranging from morality to mathematics to art – and yet, when we look around the world for the objects and properties that could make many such judgements true, we find ourselves at a loss.'

need the claim that the kind of existence is grounded in truth. To go down this road, the pluralist needs some claim like this; but I have presented a serious problem for this strategy.

### 2.2.3 Lynch: Causal Im/potence and Mind In/dependence

Michael Lynch's version of the scope problem, which is probably its most detailed exposition to date, can be viewed as filling in this argumentative gap.<sup>22</sup> Lynch does not merely appeal to a difference in kind in the "parts of reality" with which the truthbearers are concerned, but to a particular understanding of the correspondence relation. To get the argument going, Lynch assumes a particular *causal* interpretation of the correspondence theory, like the ones outlined in §1.2.2. The idea is that we analyse correspondence in terms of some causal or teleological relation such that, in order for a concept (qua subpropositional element) to correspond with some feature of the world, it must be causally responsive to that feature of the world. Now, insofar as this is merely one version of the correspondence theory, Lynch's argument is respectively weakened.<sup>23</sup> But for those sympathetic with this version, the resultant argument threatens to pose a serious problem. And those correspondence theorists who would resist the argument on these grounds must say what their alternative account of correspondence is.

Given this causal interpretation of correspondence, Lynch points out that

'where responsiveness is not plausible – either because the states in question aren't appropriately causally responsive or because the external environment contains no Gs that can be so causally responsive – then it is less likely that mental-states with G-ish content have that content because they represent Gs... [I]f we nonetheless wish to maintain that the relevant mental states are *true*, some other account of what makes them true must be pushed onto the field.' (2009: 33)

We should note first that this argument is elliptical: it moves from the claim that the *content* of the states is not to be explained in terms of responsiveness, and hence correspondence, to the claim the *truth* of the states (or their content) is not to be so explained. This relies on a truth-conditional theory of content, which is – in this book, at least – part of Lynch's overall framework (2009: 50), but which might reasonably be rejected for some domains *precisely because* it's not deemed plausible that the content of the mental states in question is to be explained in terms of representation. Setting this aside, we should also be wary of the second parenthetical clause here, that 'the external environment contains no Gs that can be so causally responsive', for the reason mentioned above. It is the second clause that is important: we do not want to rely on the *absence* of

---

<sup>22</sup> Lynch (2009: ch.2).

<sup>23</sup> Lynch (2001: 724; 2004: 385) has long held that causal im/potence drives the scope problem; but only the gloss given in Cotnoir & Edwards (2015: 118) seems to pick up on this.



the relevant objects or properties (the Gs), but on their *causal impotence*, which threatens to block the causal correspondence theory. Lynch thus needs to make the case that, in some domains, the relevant objects or properties are causally impotent.<sup>24</sup>

There certainly seem to be properties like this. For instance, gerrymandered or disjunctive properties, like the property of *being either a coffee cup or a number or Lyra Belacqua or a wrestling move* – i.e., merely “abundant” properties – presumably do not have causal relevance; there is nothing we can explain in terms of such properties. However, we might think that the correspondence theory is still appropriate for truthbearers containing predicates that ascribe such properties, since the meaning of the word or concept is plausibly to be explained in terms of the disjuncts, and as long as the meaning of *these* predicates can be explained causally, then we can understand the truth of such ascriptions derivatively. (More on this idea in §7.2.)

Lynch appeals to a different ontological difference, which is the mind-in/dependence distinction again: he claims that causal theories of representation are committed to the claim that ‘[t]rue beliefs map objects that exist and have their properties mind-independently.’ (2009: 33) But why on earth should this be so? Lynch merely states that this ‘is a consequence of the fact that representational views intend [!] for their positions to be realist.’ (*ibid.*) The mind in/dependence distinction is certainly one way of carving the anti-/realist distinction, but even if such representational theories do intend to be realist in this sense (which is debatable), the mere *intention* of being so hardly commits the theory.<sup>25</sup> (Recall also that the correspondence theory is compatible with truthmakers being mental entities, as in McTaggart’s idealist correspondence theory; if correspondence theorists do *ipso facto* intend to be realist, they roundly fail.) To support this contention, the fundamental claim must be that such objects and properties are incapable ‘of entering into at least indirect causal interaction with our minds.’ (*ibid.*)

To motivate the rejection of the correspondence theory for the moral domain, Lynch suggests that ‘it is difficult to know how wrongness – even if we grant that it is a property – can be a natural property with which we can causally interact.’ (2009: 34) Note that the word ‘natural’ has suddenly cropped up here. In the context, it is clear that Lynch must mean something like *mind-independent* by this, rather than (merely?) *joint-carving* or *sparse* (which would be “natural” in Lewis’s (1983) sense). Why be sceptical of the mind-independent existence of moral properties? In the back of Lynch’s mind, perhaps, may be Mackie-style “queerness” considerations to the effect that moral properties, in all their objective-categorical-reason-giving glory, look too weird to be a part of our naturalistic world.<sup>26</sup> (Hence, perhaps, the sudden appearance of ‘natural’.) But note that it is difficult

---

<sup>24</sup> I am unsure what Lynch means by the possibility that ‘the *states* in question aren’t appropriately causally responsive’ – this has the whiff of expressivism, but of course expressivists will deny that we are to explain the content of the relevant mental states in terms of their truth conditions – see Chapter 4.

<sup>25</sup> Similar scepticism is expressed, in passing, by Edwards (2013a: 393).

<sup>26</sup> Mackie (1977). Edwards (2011: 29, especially fn.3), for instance, suggests just this concern in this context.

to see how mere *mind-dependence* should alleviate these concerns at all. If anything, this gives us the power to “create” such “queer” properties; and that, if anything, is spookier still.

In any case, Lynch seems to be relying on the idea that moral properties might be mind-dependent. For what it’s worth, I find the idea that moral properties are mind-dependent somewhat repulsive: that wantonly torturing animals or children is wrong is not to be explained to any degree by our attitudes towards these actions.<sup>27</sup> But, in any case, it seems unlikely that mind-dependence can do the work that Lynch needs it to do, for Lynch requires mind-dependence to rule out causal relevance, and why should that be? First, note that we do not want to restrict causal relevance to merely those properties that are causally *efficacious*, at the “fundamental level” of reality, say. For then the only propositions that could correspond with reality would be those that talk about the fundamental level – those formulated in “ontologese”, say – and that is a deeply undesirable result.<sup>28</sup> The correspondence theory, if it is plausible for any propositions, is plausible for propositions concerning tables and chairs, the macroscopic objects and properties of everyday life; and if it is to be cashed out causally, then these objects and properties must be causally relevant.<sup>29</sup> And just as we appeal to such objects and properties in explanation all the time, so too do we appeal to mind-dependent properties. Take the property, for instance, of *being expensive*. Nothing would be expensive if there were no minds, but that hardly stops us from explaining why we didn’t buy something in terms of its being expensive. From whence, then, the block to causal interaction from mind-dependence?

A potential explanatory contrast between mind-in/dependent properties concerns what Wright (1992: 196-9) calls “width of cosmological role”. It *may* be that mind-dependent properties are not ‘potentially contributive to the explanation of things *other than*, or *other than via*, our being in attitudinal states which take such states of affairs [i.e., involving such properties] as object.’ (1992: 196) Such things are said to have a “narrow” cosmological role. For instance, it may be that the only explanations which cite the property of *being expensive* are explanations of our beliefs concerning which things are expensive, or which cite our beliefs about which things are expensive. (After all, if something’s expense in particular (rather than merely its cost) prevents you buying it, it is usually

---

<sup>27</sup> One might think that what makes these actions wrong is that they cause pain, and without minds there is no pain. But this only makes wrongness “mind-dependent” in the sense that pain is “mind-dependent”; as is well-known, the fact that mental states like pain fail the counterfactual criterion for mind-independence given above is a counterexample to that criterion. Even if wrongness is only instantiated in worlds where there are minds, I do not think this makes wrongness mind-dependent in any significant sense. Note that, as an expressivist, I am ultimately committed to reading “mind-in/dependence” claims as internal to ethical discourse. Blackburn has several excellent discussions of this, but see especially (1993: 172-8; 2010: 295-303). See also Köhler (2014).  
<sup>28</sup> Though recall Barnard & Horgan’s (2006, 2013) correspondence pluralism (§1.3.2): only propositions concerning the fundamental nature of reality “directly” correspond, while others have a “mediated” correspondence.

<sup>29</sup> For the general need to admit the causal relevance of macroscopic objects and properties so as to feature in “coarse-grained” explanations, see especially Jackson & Pettit (1992a), but also their (1988, 1990, 1992b). In what follows, I will assimilate to some extent talk of what is “causally relevant” with talk of what is “explanatory”; this is in tune with a broadly causal notion of explanation, but does not presuppose that *all* explanation is causal, only that the relevant cases are.

because you *believe* it's expensive; if you believed it was cheap, then its (really) being expensive would be a poor explanation.) But, again, it is difficult to see how a narrow cosmological role is sufficient to prevent a property from entering into causal interactions. If Lynch is relying on something like narrow cosmological role in his argument here, then it needs to be made much more explicit precisely what the problem is supposed to be.<sup>30</sup>

Lynch's claim is far more plausible when it comes to abstract objects like numbers; it is difficult to see how we could enter into causal interactions with *them*. But it's worth stressing that it is the abstractness that does the work here, not mind-dependence.

Nonetheless, the causal theory of correspondence would hereby fill in the gap in the above arguments from Cotnoir & Edwards and Pedersen. Even if a number's existence can be explained in terms of a truthbearer's being superwarranted (rather than its truth), we still, it seems, cannot explain its truth in terms of correspondence if correspondence is to be understood causally and abstract entities cannot enter into causal interactions.

Even conceding the case against the causal correspondence theory in the mathematical domain, however, we are left with four serious issues. The first, most important for our purposes, is that this leaves the vast majority of domains for which correspondence is supposed to be implausible – the moral, the legal, the comic, the aesthetic, the fashionable, the modal, and so on – untouched. The second is that it is worth re-emphasising the plethora of non-anti-/realist pluralist options outlined above. The pluralist needs to argue that her proposal regarding mathematics is to be preferred over all competitors (including even correspondence pluralism). The third, related, point is that we might just take the inapplicability of the causal interpretation of the correspondence theory to the mathematical domain to be a counterexample to that interpretation of the theory. Finally, even if we grant an anti-realist truth property for the mathematical domain on this basis, we still need to run the scope problem in the reverse direction, and show that this property is locally implausible for other domains.

Here, the pluralist starts on firmer ground, since there are powerful arguments against the bare global adequacy of epistemically constrained theories of truth.<sup>31</sup> But while this would be sufficient to establish disjoint pluralism, the *domain-restricted* pluralist requires the stronger *local* inadequacy of the anti-realist theory of truth *for a particular domain*. (As noted in §1.3.2, disjoint pluralism in general does not entail domain-restricted pluralism in particular.) On this front, Lynch has remarkably little to say: he merely stresses the “plausibility” of the representational theory of content for, e.g., beliefs about medium-sized dry goods (2009: 42). As pointed out above, the plausibility of one theory does not engender the implausibility of another.

---

<sup>30</sup> Note that, while it is plausible that properties like *being expensive* have narrow cosmological role, this looks less plausible for other potentially mind-dependent properties, including potentially “socially constructed” properties, like *being a woman*; see, e.g., the papers collected in Haslanger (2012).

<sup>31</sup> See, for instance, Künne's (2003: §7.3) “argument from blind spots in the field of justification”; see also e.g. Walker (2001), Williamson (1994a), Lynch (2009: 45-8).

## 2.3 A New Strategy

All in all, this is rather disappointing. By presupposing a causal version of the correspondence theory of truth for some domains, all we've really argued is that such a theory cannot explain truth in discourses that concern causally impotent abstract objects like numbers. As far as conclusions go, this is hardly ground-breaking. Most problematic, however, is that this version of the scope problem has proven, ironically enough, remarkably restricted in scope. It leaves pluralism unmotivated for the majority of domains that pluralists trot out when the scope problem comes up.

I think that the pluralists' approach to the scope problem has been misguided on two fronts. The first is that ontologically-driven versions in general are unlikely to be as persuasive as they at first appear, for the reason covered at the end of §2.2.2: diversity in what exists still presupposes their mutual existence. The second concerns the argumentative strategy. Indeed, once it is laid bare it immediately begins to look rather odd. The pluralist starts with a multiplicity of theories of truth, or candidate truth properties –  $T_1, \dots, T_n$  – and truthbearers sorted into rough-and-ready sets –  $D_1, \dots, D_n$ . She then treats theorising about the nature of truth as an exercise in matching, truth properties to domains, on the basis of plausibility.<sup>32</sup> This might be permissible as an introductory exercise, as a way of getting a sense of the strengths and weaknesses of different theories. But as a rigorous methodology it has little to recommend it.

No, I think the pluralist should and can do better, and that doing better does not involve piecemeal arguments for localised im/plausibility. I recommend a much more direct approach. Namely, that we argue that the nature of truth for truthbearers in  $D_1$  is  $T_1$ , and we provide a separate argument that the nature of truth for truthbearers in the disjoint set  $D_2$  is  $T_2$ . Unless there's something wrong with the arguments, this entails that the nature of truth varies between truthbearers. On the direct strategy, "implausibility" claims are simply redundant. My view is that we can pursue the direct strategy if we assume an underlying metasemantic pluralism.

That metasemantic and truth pluralism might come as a package was hinted at by Crispin Wright, during an exchange with Simon Blackburn:

'For Blackburn, the relevant plurality is not among ways in which truth may be constituted in different regions of discourse, but among *bearers* of truth – the realism-relevant distinctions are to be made by seeking out variety among kinds of *propositions*... That *sounds* as if it could be a real theoretical contrast. But on closer inspection there are difficulties in seeing how it could be stable... The problem is to understand what *kind* of distinctions among types of proposition, of the sort Blackburn wants to draw, might be constitutive of

---

<sup>32</sup> So the arguments for our positive proposals amount to little more than "If not this, then what?"

the realist/anti-realist contrasts yet go unreflected by systematic differences in what makes for their *truth*.’ (1998a: 191)<sup>33</sup>

But, again, the idea here is presented only suggestively: it may be that distinctions among types of proposition (more precisely: in how we explain the meaning of different sentences) get “reflected” in distinctions in the constitution of truth, but this might equally be a case of “double-counting”. It is the task of Part II to follow up on this idea, to show that metasemantic pluralism of a Blackburnian variety gives us good reason to endorse alethic pluralism of a Wrightian variety (and thereby marry the views of two of my philosophical grandfathers).<sup>34</sup>

---

<sup>33</sup> Wright also intriguingly suggests that ‘[t]he converse direction of implication is also plausible.’ (*ibid.*)

<sup>34</sup> My other grandfather is Robert Stalnaker (supervised Paolo Santorio; Wright supervised J. Robert G. Williams; Blackburn supervised Daniel Elstein; I have no philosophical grandmothers). Perhaps the next project should be to put a Stalnakerian spin on the whole thing. Ideas on a postcard.

---

PART II

A METASEMANTICALLY-DRIVEN  
ARGUMENT FOR TRUTH PLURALISM

---

---

# CHAPTER 3

## EXPLANATION, SUCCESS, AND COINCIDENCE

---

‘For we can think of “pertinence” as a psychological relation, holding between an agent’s beliefs and desires, on the one hand, and actions... on the other; or we can envisage “pertinence” as a relation between actions and outcomes... The explanatory mystery is precisely why there should be any action that meets these two constraints.’

Kitcher (2002: 359).

Deflationists endorse an anomalous quietism about the property of truth (§1.2.1). This is defended on the grounds that truth itself is anomalous: truth, deflationists contend, is explanatorily impotent. That is, we do not need to postulate it to explain any phenomena. Rather, the truth predicate is merely a device that enhances the expressive capacity of our language, and to play this expressive role, we do not need to understand the truth predicate as representing some substantive property, out there in the world, with an interesting nature that might admit of philosophical elucidation. In this chapter, I argue that the deflationist’s anti-explanatory contention is false.<sup>1</sup>

The Success Argument suggests that truth is a causal-explanatory property like any other on the grounds that *true belief facilitates practical success*: agents that believe the truth tend to be more successful than those who believe falsely. Deflationists, however, have a seductive reply that purports to show that the sole role for the truth predicate here is expressive. Truth itself, we are told, can be eliminated from the explanation of any particular instance of practical success in favour of (roughly) an instance of the deflationist’s preferred schema;<sup>2</sup> the role for the truth predicate is to generalise over these individualised explanations. I set out the argument and the deflationist’s response in Section 1.

Section 2 presents a fatal problem for this strategy. Analogues of the deflationist’s individualised explanations can be provided by way of explanation of coincidental instances of practical success where the agent has false beliefs; so, by the deflationist’s lights, there appears to be no substantive explanatory difference between coincidental and non-coincidental instances of practical success. But the non-/coincidental distinction *just is* an explanatory distinction. The deflationist, then, cannot make sense of this contrast; her explanatory strategy treats every instance

---

<sup>1</sup> The central argument appears in Gamester (fc).

<sup>2</sup> For ease of presentation, I address the argument here to the most prominent varieties of deflationism: disquotationalism and minimalism. But deflationists are united in their anti-explanatory contention.

of practical success as though it is coincidental! The inflationist, by contrast, can straightforwardly offer an account by invoking the explanatory potency of truth. In Section 3, I argue that the deflationist's prospects for recovering the contrast by extending her individualised explanations without invoking truth are, at best, bleak, and contend further that the deflationist cannot hope to make sense of the *correlation* between practical success and truth. Indeed, I suggest the deflationist's entire strategy manifests an underlying confusion about the nature of explanation.

This establishes an explanatory role for truth that I will argue in the next chapter is played by a realist truth property.

### 3.1 Deflating Success

#### 3.1.1 The Success Argument

The Success Argument for inflationism is that truth must be a causal-explanatory property like any other because *true belief facilitates practical success*: we need to postulate truth if we are to explain the practical success of certain actions performed by rational agents.<sup>3</sup> Let's illustrate this via the easiest gameshow in the world: *Be a Millionaire!* The contestant sits with two opaque boxes in front of her – box X and box Y. The host puts a cheque for a million pounds into one of the boxes. The rules dictate that all the contestant has to do is pick the box with the cheque in, and she wins the money. Both Alison and Bryan – desperate to be millionaires – go on the gameshow, but have quite different experiences:

**Alison.** The host puts the cheque in box X. Alison chooses box X. She wins the money.

**Bryan.** The host puts the cheque in box X. Unbeknownst to Bryan, the cheque is surreptitiously moved to box Y. Bryan chooses box X. He doesn't win the money.

The actions share a distinctive feature, which we will call *psychological pertinence*:<sup>4</sup>

#### *Psychological Pertinence*

An action A is psychologically pertinent to an outcome O for a rational agent S in a context C iff: (i) A is an option for action for S in C; and (ii) S believes that S's A-ing in C will result in O.

---

<sup>3</sup> For some of the relevant history, see Field (1972, 1986), Putnam (1978), Leeds (1978, 1995, 2007), Kitcher (2002), Damnjanovic (2005), and Wrenn (2011).

<sup>4</sup> The distinction between psychological and effective pertinence is due to Kitcher (2002: 358-9) and the terminology is Wrenn's (2011: 460). My own definitions elaborate substantially on their sources; there are subtle variations between the accounts, but they need not delay us here.



The basic idea is simple: for any agent in any context, the action psychologically pertinent to a particular outcome is the one that the agent believes will result in that outcome if she were to perform it. This means that, if the agent desires that O and has no countervailing desires, then the agent will perform the action psychologically pertinent to O. We can say that an action A is psychologically pertinent with respect to a desire D (for S in C) iff D is the desire that O and A is psychologically pertinent to O (for S in C). Psychological pertinence is thus a necessary condition on intentional explanation of rational action: if we cite a set of beliefs and a desire in the intentional explanation of an agent's action, then (on the assumption that the agent acted rationally) we are claiming that the action was psychologically pertinent with respect to the desire according to those beliefs.<sup>5</sup>

The slightly clunky locution 'S's A-ing in C will result in O' is left deliberately vague in order to cover the myriad different connections that an agent might imagine holds between her performing an action and an outcome coming about (via magic or the will of God, for example, as well as more humdrum causal connections, or even a giant question mark if one is unsure of why the one will bring about the other). Whatever else she believes, the agent must believe that her performance of A will be in *some* way explanatorily relevant to O coming about. For this reason also, the 'will result in' locution is preferable to simple conditional 'if S As in C then O'.

Because their relevant beliefs are effectively identical (we can stipulate) and both have the desire for wealth, the same action is psychologically pertinent with respect to the relevant desire for both Alison and Bryan. However, only Alison's action actually satisfies her desire. Thus Alison's action has a feature that Bryan's lacks, which we will call 'effective pertinence':

*Effective Pertinence*

An action A is effectively pertinent to an outcome O for an agent S in context C iff S's A-ing in C will result in O.

---

<sup>5</sup> In fact, psychological pertinence as it stands is too simple to play this role in full generality, being defined in terms of binary beliefs. We can straightforwardly define a more sophisticated notion of psychological pertinence in terms of credence that is apt to play the role more widely:

*Psychological Pertinence\**

An action  $A_i$  is psychologically pertinent\* to an outcome O for a rational agent S in a context C iff: (i)  $A_i \in \{A_1, \dots, A_n\}$ , where  $A_1, \dots, A_n$  are S's options for action in C; and (ii) for any  $j \neq i$ , where  $A_j \in \{A_1, \dots, A_n\}$ , the credence that S assigns to S's  $A_j$ -ing resulting in O is less than or equal to the credence S assigns to S's  $A_i$ -ing resulting in O.

While a more adequate rationalisation condition, talking in terms of psychological pertinence\* (and the relevantly altered notion of effective pertinence) results in unnecessary complication. All that we require are the simple cases in which (non-starred) psychological pertinence is satisfactory.

That is, an action is effectively pertinent to an outcome if performing that action will *in fact* result in that outcome. An action A is therefore effectively pertinent with respect to S's desire D (in C) iff D is the desire that O and A is effectively pertinent to O (for S in C).

Note that, for an action to be effectively pertinent with respect to an agent's desire, it need not be the case that the agent performs the action at all – and even if they do, it is not necessary that they perform the action *because* they have that desire. For example, drinking this coffee will mean that I am more alert for the upcoming seminar (let's say), which I certainly want to be. The action itself is therefore effectively pertinent with respect to this desire even if I (i) don't drink the coffee; (ii) drink the coffee despite being unaware of its caffeine-based benefits, because I like the taste (for example); or (iii) have the relevant belief about these benefits, but nonetheless drink the coffee for a different reason, e.g., whilst under duress.

Nonetheless, with effective pertinence in hand it is tempting to introduce a definition of *success*, understood as a species of effective pertinence:

*Success*

An action A is successful with respect to a desire D for an agent S in a context C iff: (i) S's A-ing in C is effectively pertinent with respect to D; and (ii) D correctly features in the intentional explanation of S's A-ing in C.

The idea is just that an action is successful iff it satisfies an agent's desire, and the agent performed that action *in order to* satisfy that desire.<sup>6</sup> Note that actions are successful *with respect to desires*; this is because it seems implausible that it will always be the case that exactly one desire correctly features in the intentional explanation of an action: beer drinkers will often insist that they drink beer not only because they want to get drunk, but *also* because they like the taste. Furthermore, it's entirely possible that such an action will be effectively pertinent with respect to one of these desires and not the other, e.g., if the beer tastes nice, but is secretly non-alcoholic (or, more usually, is gross but gets you drunk). Perhaps an action can be said to be successful *simpliciter* iff it is successful with respect to every desire that correctly features in its intentional explanation – we have no direct need for such a notion.

Given that any action performed by a rational agent will be psychologically pertinent with respect to the desire(s) that correctly feature in its explanation, if an action A is successful with respect to a desire D (for S in C) then the psychological and effective pertinence of A with respect to D coincide (for S in C).<sup>7</sup> Nothing hangs on this explication of 'success' being the "correct analysis" of this concept, if such there be – if one has quibbles about cases, that should not defeat our purpose. All that the argument strictly requires is this coincidence of psychological with effective pertinence

---

<sup>6</sup> Compare Wrenn (2011: 454): '*success* occurs when a particular action has its intended outcome.'

<sup>7</sup> Note that the success of A with respect to D is only sufficient for the coincidence of psychological and effective pertinence, not necessary.

of an action with respect to a desire that features in the explanation of that action: the agent performing an action because she thinks it will result in a particular outcome, and the action actually resulting in that very outcome.

Alison’s action is not only psychologically pertinent with respect to her desire for wealth, but also effectively pertinent with respect to this desire; indeed, it is *successful* with respect to this desire. Bryan’s is not. And this despite the fact that both agents have equivalent, equally well-justified beliefs with respect to their present situation, and thus which action would result in the satisfaction of their desire for wealth. The sole difference between the cases is that Alison’s belief about the location of the cheque was true while Bryan’s was false. Thus the inflationist about truth *wants* to say that truth plays a role in explaining why Alison’s action turned out to be effectively pertinent: Alison’s action turned out to be successful *because Alison’s beliefs were true*. In particular, the beliefs that feature in the intentional explanation of the action – those that explain the psychological pertinence of the action – were true.

The onus is on the deflationist at this point. Actions that result from true beliefs (via psychological pertinence) tend to be effectively pertinent with respect to the desires that feature in their intentional explanations along with these beliefs – that is, they tend to be successful. This seems like a prime explanatory role for truth. The deflationist thus owes us an account of how we can possibly explain the success of Alison’s action *without* appealing to the truth of her beliefs.

### 3.1.2 Individualised Explanations

The deflationist, however, has a seductive response.<sup>8</sup> She does not deny that true beliefs facilitate practical success; instead, the strategy is to *earn the right* to this claim on deflationist-friendly grounds, by arguing that the sole role for the truth predicate here is expressive; in particular, that it functions as a device for generalisation. To this end, we are told that, strictly speaking, any mention of ‘truth’ can be eliminated from the explanation of each *individual* instance of practical success in favour of the relevant instance of the deflationist’s preferred schema. The truth predicate merely exploits this structural similarity, and thereby generalises over these individualised explanations.

The deflationist asks us to consider any individual instance of practical success. Horwich’s (1998: 22-3) example is of an agent, Bill, who desires a beer and believes he can get one simply by nodding – which, in fact, he can. The deflationist’s explanation of Bill’s success is as follows:

- |     |  |                |
|-----|--|----------------|
| (1) | Bill desires that Bill has a beer.                                   | <i>premise</i> |
| (2) | Bill believes that Bill’s nodding will result in Bill having a beer. | <i>premise</i> |

---

<sup>8</sup> Damjanovic describes the following as the deflationist’s ‘standard response’.

- (3) For any rational agent S and option for action A, if S believes that S's A-ing will result in O and S desires that O, then (other things being equal) S will A. *premise*
- (4) Bill is a rational agent. *premise*
- (5) If Bill believes that Bill's nodding will result in Bill having a beer and Bill desires that Bill has a beer, then (other things being equal) Bill will nod. *from (3), (4)*
- (6) Bill nods. *from (1), (2), (5)*
- (7) Bill's nodding will result in Bill having a beer. *premise*
- (8) Bill has a beer. *from (6), (7)*
- (9) Bill's desire that Bill has a beer is satisfied. *from (1), (8), def. of satisfaction*

I have tried to make the deflationist's argument as explicit as possible. (1) and (2) are treated as premises, although they themselves could in principle be explained, presumably via Bill's psychological history. The content of the belief in (2) is usually given in conditional form:

- (2\*) Bill believes that: if Bill nods, then Bill has a beer.

The same goes for (7). Once again, the 'will result in' locution allows us to make explicit that Bill must believe that there's an explanatory connection between his nodding and his getting a beer; and, in (7), there must actually *be* such a connection.

We are normally presented with just (5), rather than (3), (4), and (5). However, for (5) to truly be explanatory it cannot simply be a *sui generis* fact about Bill that when he has that belief and desire he will perform that action. Rather, it must be an instance of a more general psychological phenomenon (and it is clear that this is how Horwich intends for it to be understood). This is given by the "practical syllogism" of (3), along with the assumption that Bill is a rational agent (4). I have also made explicit the *ceteris paribus* clause. Note that (3) essentially tells us that, if an action is psychologically pertinent with respect to an agent's desire (for that agent in that context), then other things being equal she will perform that action.

Given the *ceteris paribus* clause, the inference from (1), (2), and (5) to (6) is not deductive, but as we are treating all things as being equal, we can simplify things by treating it as such.

(7) is also a premise of the case, but it's important to emphasise that (7), like (5), cannot be explanatory if it is somehow a *sui generis* fact about Bill's nodding – it must itself be explicable. Unlike in the case of (5), there is no single obvious candidate for this explanation, and the right explanation will depend entirely on how we flesh out the details of the case. For instance, it might be that Bill is in a bar where the bartender is aware of Bill's introverted nature and drink preference, and thus will

give Bill a beer when he nods; or perhaps Bill has attached his head to a Wallace-and-Gromit-style beer-dispensing device, such that when he nods he gets a beer. The details do not matter; but that there *are* such details does.

The rest of the inference requires no further elaboration.

### 3.1.3 Truth's Expressive Role

The inference given in the argument from (1) to (9) is the lynchpin of the deflationist's response to the Success Argument. The deflationist hereby provides what we might call a "two-step" account of Bill's success. We start with *an explanation of why Bill performed a certain action* in (1) through (6). This explanation goes via the psychological pertinence of the action – given in (2) – with respect to a desired outcome (1). Since we know that Bill, qua rational agent, will (other things being equal) perform actions psychologically pertinent to his desires (as stated in (5), obtained via (3) and (4)), Bill's propositional attitudes hereby explain why Bill nodded (6). The second step involves *an explanation of why this action ended up satisfying Bill's desire* in (6) through (9). This explanation goes via the effective pertinence – given in (7) – of the action (6) to the desired outcome (8). *Together*, the two steps are supposed to provide a full explanation of the *success* of Bill's action. Bill performed the action because it was psychologically pertinent to an outcome he desired, and the action resulted in the satisfaction of this desire because it was, in fact, effectively pertinent to this outcome.

The critical premises are (2) and (7):

- (2) Bill believes that Bill's nodding will result in Bill having a beer.
- (7) Bill's nodding will result in Bill having a beer.

There are at least three important features to note.

First, (2) effectively states the *psychological pertinence* of Bill's nodding with respect to his desire for a beer (for Bill in the present context); while (7) states the *effective pertinence* of this action with respect to this desire. Second – as noted above – while (2) and (7) are included as premises, they can in principle be explained themselves; that is, we can *explain* the psychological and effective pertinence of the action. This will be important later.

Most significant in the present context is that (2) and (7) together constitute an instance of the schema 'S believes that p and p'. For the deflationist, the "truth" of Bill's belief consists in nothing more than the relevant instance of this schema being so. So, the "truth" of Bill's belief *does* make an appearance in the explanation of Bill's success, but this requires nothing more than is given by the relevant instance of the schema. The truth predicate then exploits this structural similarity to generalise over the true beliefs. By the very nature of the schema, the truth of an "action-guiding" belief like (2) will *consist in* the psychological and effective pertinence of the relevant action with

respect to the relevant desire.<sup>9</sup> And this is what gives the deflationist’s purported explanation its seductive charm, for given the psychological and effective pertinence of the action, then as long as the agent is rational, the resultant action will be successful. Furthermore, it seems as though this style of explanation is entirely generalisable: after all, for any successful rational action there will be a psychological explanation of why the agent performed that action and a “worldly” explanation of why it resulted in the satisfaction of the desire. Therefore, the deflationist argues, when we say that true belief facilitates practical success, the role of the truth predicate is simply one of generalisation: to generalise over these particularised explanations.

The deflationist’s response has not been without its critics; but the argument is incredibly seductive precisely because it is difficult to see what could be *missing* from the deflationist’s (1)-(9) explanation of Bill’s success. There simply seems to be no explanatory work left for an inflationary truth property to do; so it becomes very tempting to think that truth must merely be generalising over these individualised explanations.<sup>10</sup> My goal in the rest of this paper is to argue directly that these individualised explanation are, in fact, problematically incomplete.

## 3.2 A Problem Case

### 3.2.1 *Coincidental Success*

Let us introduce our problem case; a third contestant, Callum:

**Callum.** The host puts the cheque in box X. Unbeknownst to Callum, the cheque is surreptitiously moved to box Y. Callum chooses box X. In the audience is an eccentric billionaire, who loves box X. She is so pleased that Callum has chosen her favourite box that she gifts him a million pounds.

Callum is like both Alison and Bryan in that his action is psychologically pertinent with respect to his desire for wealth. He is like Bryan, but *unlike* Alison, in that his belief about the location

---

<sup>9</sup> We are granting the deflationist the view that such beliefs appear in the intentional explanations of rational actions; Wrenn (2011) contends that this “single action-guiding belief” view of intentional explanation is controversial.

<sup>10</sup> The *style* of this response to the Success Argument exemplifies a powerful dialectical move in any dispute between inflationists and deflationists. For any purported explanatory or normative role for truth, the deflationist seeks to eliminate mention of truth from any particular instance of the explanation or norm in favour of the relevant instance of their preferred schema; and thereby maintain that the role of the truth predicate is merely as an expressive device for indirect reference to and compendious generalisation over such instances. Much of Horwich’s (1998) is occupied with applying precisely this strategy to various inflationary arguments. Another particularly neat example is Dodd’s (1999) response to Wright’s (1992) argument that truth marks a distinctive norm of assertion: “[‘One should assert only what is true’] is just shorthand for an infinite conjunction of claims of the form... [‘One should assert that snow is white only if snow is white; one should assert that coal is black only if coal is black, etc.’]” (1999: 296).

of the cheque (which explains the psychological pertinence) is false. But he is like Alison, and *unlike* Bryan, in that his action *is* effectively pertinent – indeed, *successful* – with respect to his desire for wealth.

Despite this alignment of psychological and effective pertinence, there is a salient difference between Callum’s success and Alison’s success. While Alison’s success is to be expected, Callum’s is quite out of the blue – the fact that the action he *thought* would result in the satisfaction of his desire *in fact* turned out to do so is a *sheer coincidence*.

(To be clear, there is nothing coincidental or inexplicable about *the fact that Callum performed a certain action*; nor is there anything coincidental or inexplicable about *the fact that Callum’s action had the effects it had*, in and of themselves. Each of these can be explained. It is Callum’s *success* that is coincidental, i.e., that *the same* action turned out to be *both* psychologically and effectively pertinent for Callum. Alison’s success, by contrast, is a paradigmatic non-coincidence.)

There is nothing particularly special about the existence of coincidental instances of practical success in and of themselves. The non-/coincidental contrast is ubiquitous. It is probably a coincidence that the Sun and the Moon appear to be roughly the same size from the Earth; though it might be the result of intentional design. Two people may look alike by sheer coincidence; or they might be relatives. I may have pressed my hands into some wet concrete and thereby left an imprint into which my hands perfectly fit; or the rain may have eroded that shape into the concrete by chance, making the perfect fit sheer coincidence.

The significance of the contrast for our purposes is that it is often treated as a truism that coincidences *cannot be explained*.<sup>11</sup> In fact, my argument only requires the weaker, and necessarily more plausible, claim that there is a difference in kind between what *can* be said by way of explanation when something is a coincidence, and when it is a non-coincidence. For of course there is *something* we can say by way of explanation when something is a coincidence. If it is a coincidence that the Sun and the Moon appear to be the same size, then the best we can do is offer an explanation of why the Sun appears to be that size (in terms of its size and distance from Earth) and a distinct explanation of why the Moon appears to be that size (in terms of *its* size and distance from Earth), and conjoin them together. That is, the best we can do is conjoin two *distinct* explanations of each of the relevant incidences that make up the *coincidence*. This quasi-explanatory pattern is pervasive. *If it is a coincidence* that two people look alike, all we can offer is an explanation of why one person looks the way they do (in terms of genes and environment) conjoined with an explanation of why the other person looks the way they do (in terms of genes and environment). *If it is only a coincidence*

---

<sup>11</sup> A recent and very insightful discussion of coincidence starts with the sentence, ‘It is a truism that coincidences cannot be explained’ – see Lando (2017) and the citations therein. Lando suggests that this truism ought to be finessed, because there are coincidences with a common cause; in particular, she suggests coincidences may be cases where there is a salient *relational property* between the instances that cannot be explained (e.g., that the Sun and Moon appear *the same* size; that two people look *alike*; and so on). I like this suggestion, and it accords well with what I say below (in Callum’s case, it is a coincidence that the psychologically and effectively pertinent action are *the same*); but I need not rely on it.

that there is an imprint in the concrete that my hands fit perfectly into, then all we can offer is an explanation for why my hands are a certain shape (in terms of genes and environment) and an explanation of why the imprint is a certain shape (in terms of erosion). Since all it is for there to be a perfect fit between two things is for one to have one shape and the other to have another, the conjunction of these distinct explanations is all that can be said by way of explanation, *if the fit is merely coincidental*.

We can call these *conjunctive quasi-explanations*. In line with the aforementioned truism, it does not seem especially plausible to me that such conjunctive quasi-explanations really deserve to be called ‘explanations’, so perhaps ‘pseudo-explanations’ would be better; but we’ll use ‘quasi’ in the interests of neutrality.<sup>12</sup> All that matters for our purposes is there is an obvious difference in kind such conjunctive quasi-explanations (be they explanations or not) and explanation proper;<sup>13</sup> so the kind of conjunctive quasi-explanation that may be tolerable qua explanation when something is a coincidence is manifestly inadequate when it is not a coincidence. For example, if asked why two people look alike we *merely* explain why each looks the way they do, and leave out the fact that they are related, then we have portrayed what is perfectly explicable as a coincidence, and thereby failed to explain why it is they look alike. Similarly if we explain why each of the Sun and Moon appear to be the size they do in terms of their size and distance from Earth, without mentioning the intentions of a creator. To offer a conjunctive quasi-explanation of a non-coincidence is a straightforward explanatory failing. In particular, until the quasi-explanation is *supplemented* with something that *actually explains* the non-coincidence, it remains unexplained, precisely because we hereby portray what is perfectly explicable as though it is coincidental.

Bearing this explanatory contrast in mind, my contention is that the deflationist’s refusal to concede the explanatory potency of truth means she does not have the resources to distinguish the non-coincidental instances of practical success from the coincidental.

### 3.2.2 *The Threat of Collapse*

Recall Alison, our successful gameshow contestant. By the deflationist’s lights we can give a complete two-step explanation of Alison’s success via the following derivation:

- (1A) Alison desires that Alison has a million pounds. *premise*
- (2A) Alison believes that Alison’s choosing box X will result in Alison *premise*  
having a million pounds.

---

<sup>12</sup> Whether or not conjunctive quasi-explanations are really explanations, moving forward I’ll largely use ‘explanation’ to mean the kind of explanation proper that is appropriate for non-coincidences in particular.

<sup>13</sup> It is, of course, controversial exactly what an explanation proper consists in. We can remain neutral on such issues.



- (3A) For any rational agent S and option for action A, if S believes that S's A-ing will result in O and S desires that O, then (other things being equal) S will A. *premise*
- (4A) Alison is a rational agent. *premise*
- (5A) If Alison believes that Alison's choosing box X will result in Alison having a million pounds and Alison desires that Alison has a million pounds, then (other things being equal) Alison will choose box X. *from (3A), (4A)*
- (6A) Alison chooses box X. *from (1A), (2A), (5A)*
- (7A) Alison's choosing box X will result in Alison having a million pounds. *premise*
- (8A) Alison has a million pounds. *from (6A), (7A)*
- (9A) Alison's desire that Alison has a million pounds is satisfied. *from (1A), (8A), def. of satisfaction*

(1A)-(6A) explains why Alison performed a particular action, and (6A)-(9A) explains why that action resulted in the satisfaction of her desire. In Bryan's case, we can provide the first step of the explanation, but not the second.

The *problem* is that we can apply the deflationist's explanatory schema, without alteration, to Callum's case:

- (1C) Callum desires that Callum has a million pounds. *premise*
- (2C) Callum believes that Callum's choosing box X will result in Callum having a million pounds. *premise*
- (3C) For any rational agent S and option for action A, if S believes that S's A-ing will result in O and S desires that O, then (other things being equal) S will A. *premise*
- (4C) Callum is a rational agent. *premise*
- (5C) If Callum believes that Callum's choosing box X will result in Callum having a million pounds and Callum desires that Callum has a million pounds, then (other things being equal) Callum will choose box X. *from (3C), (4C)*
- (6C) Callum chooses box X. *from (1C), (2C), (5C)*
- (7C) Callum's choosing box X will result in Callum having a million pounds. *premise*
- (8C) Callum has a million pounds. *from (6C), (7C)*

- (9C) Callum's desire that Callum has a million pounds is satisfied. *from (1C), (8C), def. of satisfaction*

It is striking that (1C)-(9C) may well be an acceptable conjunctive quasi-explanation of Callum's success. Its familiar two-step form explains why he performed a certain action, and why that action resulted in the satisfaction of his desire. This is presumably all there is to say by way of explanation *when someone's success is coincidental*.

But Alison's success is not coincidental. As such, there must be some explanatory contrast between Alison and Callum; something that *explains why* Alison's action was successful, but which cannot explain why Callum's action was so. The question is how the deflationist can make sense of this contrast. If *all* there is to the explanation of Alison's success is (1A)-(9A), just as *all* there is to the explanation of Callum's success is (1C)-(9C), then there is no such contrast: the two are on an explanatory footing. If this is right, then we lose all right to say that there is any non-/coincidental contrast between the cases; and this is a straightforwardly disastrous result. In my view, the deflationist is hereby exposed as treating every instance of practical success as though it is a coincidence. At the very least, what this shows is that we must go *beyond* the (1)-(9) schema if we are to truly *explain* the non-coincidental instances of practical success, and thereby recover the non-/coincidental contrast.

Furthermore, it is simply obvious what the *inflationist* will say at this point. The inflationist will appeal to some explanatory premise that invokes truth: that actions that result from true beliefs tend to be successful (i.e., effectively pertinent with regards to the desires that they result from), for example. Since the beliefs that Alison's action resulted from are true, this *explains* the success of Alison's action;<sup>14</sup> and since the beliefs that Callum's action resulted from are *not* true, it cannot do so for Callum. This is exactly the kind of explanatory tie-breaker we need. Moreover, it's intuitively very appealing: it simply seems obvious that what explains Alison's success is that she has true beliefs, while Callum is successful despite having false beliefs. This is why Callum's success is merely coincidental while Alison's is straightforwardly explicable.

This move, however, is not open to the deflationist, for the deflationist's goal is precisely to *deny* this kind of explanatory role for truth. The deflationist is thus stuck with the unenviable task of trying to explain Alison's success *in particular*, by reference to something that is true of Alison but not true of Callum *other than* the truth of her beliefs. There are broadly two strategies here, both of which we shall consider: either trying to add an additional explanatory premise, as the inflationist

---

<sup>14</sup> Note that (1A)-(9A) may still be a *proper part* of the explanation of Alison's success; as indicated above, there need not be anything strictly speaking *incorrect* about a conjunctive quasi-explanation when the explanandum is non-coincidental. It is rather that it is problematically *incomplete* (in the sense that there are explanatory premises missing), in such a way that what we have is not really an *explanation* of the non-coincidence until it is supplemented.

does, but which does not appeal to truth; or by extending the schema by looking to the explanations of the explanatory premises used.

But note that, *whatever* the deflationist goes on to say at this point, this is already a major dialectical victory for the inflationist. The deflationist’s goal was to show that truth is merely a device for generalisation in the principle that true belief facilitates practical success by showing that each individual instance of practical success can be explained without reference to truth. The argument was seductive precisely because the (1)-(9) “explanations” seemed to leave no explanatory work for an inflationary truth property to do. But we’ve now seen that the (1)-(9) “explanations” *must* be incomplete, for otherwise we risk collapsing the distinction between coincidental and non-coincidental instances of practical success. And we’ve also seen that the inflationist is entitled to the obvious additional explanatory resources required to recover this contrast. (It is difficult to see what, other than antecedent deflationary commitments, could lead us to resist it.) As far as the deflationist is concerned, then, we are back to square one: she now has to show that we can explain any *non-coincidental* instance of practical success without reference to truth; and this is just a slightly finessed version of the original challenge set out by the Success Argument. Until this is done, she cannot maintain that the sole role for the truth predicate is as a device for generalisation.

### 3.3 Twisting the Knife

#### 3.3.1 *Supplementing the Explanation?: Explaining the Premises*

Perhaps the obvious strategy for the deflationist to pursue is to try and recover the contrast between the cases by looking to the explanations of the explanatory premises used in the (1)-(9) schema. Note that, for this to work, we need more than mere *difference* in explanatory detail; any number of details may vary between what can be said by way of explanation of two instances of a phenomenon without this rendering one a mere coincidence.

In the non-coincidental case, Alison’s, we have the explanatory premise (2A), which states the psychological pertinence of Alison’s action:

- (2A) Alison believes that Alison choosing box X will result in Alison having a million pounds.

That Alison has this belief will be explained using premises like the following:

- (2Ai) Alison believes that Alison choosing the box with the cheque in will result in Alison having a million pounds.  
 (2Aii) Alison believes that the box with the cheque in is box X.

Along with some general principle about how sensible agents like Alison tend to reason given simple premises like these. Meanwhile, our coincidental case, Callum, has the following explanatory premise:

- (2C) Callum believes that Callum choosing box X will result in Callum having a million pounds.

That Callum has this belief can equally be explained by analogous premises such as the following:

- (2Ci) Callum believes that Callum choosing the box with the cheque in will result in Callum having a million pounds.  
(2Cii) Callum believes that the box with the cheque in is box X.

Along with the same general principle. As such, it is important to note that, in each case, *that the agent has the relevant belief* is perfectly explicable; and, given this, it is equally explicable in each case *why the agent performed the relevant action*. The psychological pertinence of each agent's action is, as we should expect, on an explanatory footing.

Nonetheless, there is an explanatory contrast here. The kind of simple reasoning that agents like Alison and Callum go in for given simple premises like these is valid. In Alison's case, beliefs (2Ai) and (2Aii) are both true. Therefore her resultant belief, (2A), is guaranteed to be true. By contrast, Callum's belief (2Cii) is false. There is no general principle that valid reasoning from false premises will result in a true conclusion; there is thus no guarantee that Callum's resultant belief, (2C), will be true. And yet it is. *That Callum's belief is true* is a coincidence.

Assuming that the deflationist is entitled to an adequate notion of validity, she can make sense of this explanatory contrast. In Alison's case, she can appeal to the general principle that conclusions drawn by valid inference from true premises will be true to *explain* the truth of Alison's belief. There is no such explanation of the truth of Callum's belief.

Unfortunately, as far as the present argument is concerned, this explanatory contrast is of no use *to the deflationist*. Establishing an explanatory contrast with regards to *the truth of the agents' beliefs* does not automatically establish an explanatory contrast with regards to *the success of the agents' actions*. So while the deflationist might, in this way, explain the coincidental status of *the truth of Callum's belief*, we can then ask what relevance this has to the coincidental status of *Callum's success*. Here's one way in which it might be relevant: there may be a general principle that rational actions that result from true beliefs tend to be effectively pertinent with respect to the desires that cause them. We can appeal to this principle to explain the effective pertinence (and hence success) of Alison's action; not

so with Callum. If this is right, then the fact that Callum’s belief (2C) was true by coincidence would clearly be relevant to why his action was successful by coincidence. But this is precisely the explanatory role for truth that the deflationist is denying. Indeed, it is precisely the *truth* of Alison’s belief (2Aii) that the *inflationist* thinks its explanatorily important.

Without such a principle, the deflationist cannot hope to attach any explanatory significance to the observation that Callum’s belief (2C) is true by coincidence. There are any number of features that Callum’s belief might have by coincidence: he might have had it at his favourite time of day, or just as someone else formed the exact same thought. Most of these features are not explanatorily relevant to Callum’s success; and, according to the deflationist, neither is the truth of his belief. So while this is obviously where the relevant explanatory contrast lies, the deflationist is not entitled to it, precisely because truth is purportedly *not* explanatorily relevant to practical success. Noting this explanatory contrast thus constitutes no progress at all, by deflationary lights.<sup>15</sup>

Moreover, since we can stipulate that Alison and Callum are otherwise identical (they may have identical life histories up to the point that the cheque’s location is switched), there will be no *further* difference in the explanation of the psychological pertinence of each agent’s action; after all, each agent believes that the action will result in the satisfaction of their desire *for exactly the same reasons*. By deflationary lights, then, there is nothing that might explain Alison’s success, and not Callum’s, in the explanation of the psychological pertinence of their actions.

The non-coincidental case, Alison’s, also uses premise (7A), which states the effective pertinence of the action:

(7A) Alison choosing box X will result in Alison having a million pounds.

This will be explained by premises such as the following:

(7Ai) Alison choosing the box with the cheque in will result in Alison having a million pounds.

(7Aii) The box with the cheque is in box X.

By contrast the analogous premise in the coincidental case, (7C), will permit of an entirely different explanation:

---

<sup>15</sup> The deflationist also cannot appeal to *knowledge* here, for pretty much the same reason. The reader will no doubt have noticed that Callum’s belief (2C) is a Gettier case, while Alison’s (2A) is an item of knowledge. But whatever the elusive fourth condition on knowledge might be in general, here its presence/absence is grounded in Alison’s belief (2Cii) being true while Callum’s analogous belief is false. (To reiterate once more: by design this is the only difference between the cases.) So, if the truth of Alison’s belief is not explanatorily relevant, then the knowledge/JTB contrast can’t be explanatorily relevant either.

- (7C) Callum choosing box X will result in Callum having a million pounds.
- (7Ci) Callum choosing the billionaire's favourite box will result in Callum having a million pounds.
- (7Cii) The billionaire's favourite box is box X.

*That the agent's action resulted in the agent getting the money* is nonetheless perfectly explicable in both cases, as we should expect.

There is, of course, a radical difference in explanatory detail: one mentions the location of the cheque, the other an eccentric billionaire. But this too is of no use to the deflationist. For, as mentioned, mere difference in explanatory detail is insufficient to render one instance a coincidence and the other not. If a successful contestant's cheque had been placed in box Y instead of box X, then the explanation of the effective pertinence of choosing that box would be *different* (i.e., mentioning box Y rather than box X), but her success would not be coincidental for that. What is it about *these* facts featuring in the explanation of the effective pertinence of Callum's action that is supposed to render his success coincidental? After all, if Callum had *believed* that the eccentric billionaire would act in this way, and had chosen box X on this basis, then his success would have been non-coincidental. So, it does not seem to be anything about these facts *in and of themselves* explaining the effective pertinence of the action that makes for an explanatory contrast.

The obvious *structural* difference here is that the beliefs that explain the psychological pertinence of Alison's action (2Ai) and (2Aii) *match* the worldly facts that explain the effective pertinence of her action (7Ai) and (7Aii) – her action is successful for precisely the reasons she thinks it will be. Not so for Callum. And it is obvious that this is what we *want* to appeal to in accounting for the non-/coincidental contrast between the cases. We *want* to say that an agent's actions tend to be successful when there is this matching between her beliefs and the world; and so to *explain* Alison's success with reference to this matching. But to appeal to such a relation between Alison's beliefs and the world to explain her success is to concede the game to the (realist) inflationist, for it is to postulate an explanatorily potent relation! The explanatory significance of this matching is *precisely what the deflationist is denying*.

The mere explanation of the effective pertinence of Alison's and Callum's actions has thus not turned up anything of explanatory relevance to the practical success of their actions by deflationary lights either. This is unsurprising. For, as emphasised, *that the agent chose box X* is perfectly explicable in each case; as is the fact that *choosing box X made them rich*. What is surprising in Callum's case is that *the same* action is *both* psychologically and effectively pertinent; and the natural account of this is that there is a relation that holds between the beliefs that explain the psychological pertinence of Alison's action and the facts that explain its effective pertinence that does not hold in Callum's case. But the deflationist refuses to explain Alison's success by reference to this relation.

It is difficult to see how there could be any relevant explanatory contrast in the explanation of any of the *other* premises, besides (2A)/(2C) and (7A)/(7C). I therefore cannot see how the deflationist can hope to recover the non-/coincidental contrast in the explanation of the premises used in the (1)-(9) inferences. At the very least, this leaves the deflationist with all the work to do, while the inflationary move is obvious.

### 3.3.2 *Supplementing the Explanation?: Further Premises*

It is worth being clear on the nature of the challenge here. The point is *not* supposed to be that a deflationary truth predicate would be unable to *generalise over* the non-coincidental instances of practical success in particular. The point is just the opposite: the truth generalisation is not *just* a generalisation; it is a *selective* generalisation. In particular, it excludes *coincidental* instances of practical success, like Callum's, where the agent is merely lucky enough to have the right false beliefs. Instead, it generalises over those *non-coincidental* cases of practical success, like Alison's, where the agent has true beliefs. The question is: what sense can the deflationist make of the non-/coincidental contrast between these cases, given the proposed explanatory redundancy of truth?

The deflationist proposes that the explanatory potency of "truth" is merely the explanatory potency of the psychological fact (S believes that p) and "worldly" fact (p) that makes up the relevant instance of the schema. In particular, there is no explanatorily potent relation that holds between the psychological and worldly fact when they accord with the schema. Accordance with the schema is thus not explanatorily significant, but only expressively significant: it allows us to use the truth predicate to generalise over these cases. The difficulty is that we can *also* provide a conjunction of psychological and worldly facts in coincidental cases like Callum's; just not ones that accord with the schema. All this seems to entail, by deflationary lights, is that the truth predicate cannot be used to generalise over such cases: by the deflationist's own lights, this is an *expressive* difference, not an *explanatory* one. If the deflationist maintains that *according with the schema* has some explanatory import that *not according with the schema* does not, then she is conceding the game to the inflationist; for then truth is making an explanatory contribution after all.

Grant, then, that the deflationist can generalise over the non-coincidental instances of practical success; the question is what relevance being able to so generalise has to the non-/coincidental contrast between the cases. Given that this is an explanatory contrast, if the deflationist is right that truth is explanatorily impotent, the answer must be: no relevance whatsoever. Indeed, this constitutes the problem for the deflationist, not a solution: why should it be that the cases generalised over using an explanatorily impotent predicate are of a different explanatory status to those excluded by the generalisation?

We have seen that the prospects for recovering the non-/coincidental contrast by explaining the premises used in each (1)-(9) inference are bleak. The other option for the deflationist, then, is

to add some *further* explanatory premise that explains Alison’s success in particular, but which does not make reference to the truth of her beliefs.

I have said that the inflationist will appeal to an additional explanatory generalisation, along the lines of (EG), to explain Alison’s success:

(EG) Actions that result from true beliefs tend to be successful.

A natural line of response for the deflationist is to try and utilise a schematic thesis where the inflationist seeks to use an explanatory generalisation. In this way, the deflationist may hope to explain the success of Alison’s action on deflationist-friendly grounds. It will be helpful to consider this line of resistance and where it goes wrong. First, it is not clear exactly what schema the deflationist should try and use here. (S1) will not do:

(S1) If an action results from the belief that  $p$  and  $p$ , then it will tend to be successful.

Not many *single* beliefs can be as useful as (S1) makes out. The deflationist will require something more like:

(S2) If an action results from the belief that  $p$  and the belief that  $q$  and ..., and  $p$  and  $q$  and ..., then it will tend to be successful.

Now, the thought might be as follows. Given some such schema as this, the deflationist can rearticulate it using a deflationary truth predicate as: ‘Actions that result from true beliefs tend to be successful’; or, in other words, as something very close to precisely the generalisation the inflationist uses as an explanatory tie-breaker. Given this – the reasoning might go – she is as entitled to the non-/coincidental contrast as the inflationist.

However, even if the deflationist can articulate some such principle using some such schema, a moment’s reflection reveals that the deflationist cannot use this as an explanatory tie-breaker, as the inflationist does. For recall the deflationary strategy: to *earn the right* to the claim that true belief facilitates practical success by *first* explaining (or providing a recipe for explaining) each individual instance of practical success, and *then* generalising over them. It is only in this way that she can maintain that the sole role for truth is as a device for generalisation. So *for the deflationist* the generalisation necessarily comes later in the explanatory order; she thus cannot appeal to it in the explanation of individual instances of practical success, as the inflationist does (EG). It is also worth explicitly mentioning that (S2) *itself* cannot be invoked in the explanation of any instance of practical success; for – lest we forget – (S2) is a *schema*. It is only the *instances* of a schema that have content.



To reiterate the point from above, it may well be the case that (S2) (or some schematic thesis like it) has many true instances; that these instances include the relevant non-coincidental instances of practical success; and thus that the deflationary truth predicate can generalise over these instances. It is worth reiterating once more that the mere fact that certain non-coincidental cases accord with a schema does not (by deflationary lights) *explain* why they are non-coincidental; it *merely* allows them to be generalised over. There are certainly coincidental cases, like Callum's, that do *not* accord with the schema. But the challenge to the deflationist is to make sense of the fact that these cases are of a different explanatory status (being coincidental), without conceding that the schema latches onto an explanatorily potent property.

The deflationist therefore cannot use a schematic thesis like (S2) (or a rearticulation of it formulated via a deflationary truth predicate) to break the explanatory deadlock here. And, by design, there is very little *else* that differs between Alison and Callum, and thus very little else which might be invoked to establish an explanatory contrast between the two. While the deflationist is able to generalise over the non-coincidental cases of practical success and exclude the coincidental cases, she is unable to say what this explanatory contrast *consists in*.

### 3.3.3 *The Inadequacy of Mere Generalisation*

I have argued that the deflationist's prospects for making sense of the non-/coincidental contrast between Alison and Callum are pretty miserable, but let me finish by suggesting that the deflationist's entire approach, for all its seductive charm, is *fundamentally* flawed; that it manifests an underlying confusion about the nature of explanation. The argument here proceeds via consideration of what the deflationist can say about the *correlation* between truth and practical success.

The deflationist is keen to emphasise the distinction between explanation of a phenomenon and what we might call *mere generalisation* over particular instances of the phenomenon. Very roughly speaking, in the former case a generalisation is formed by referring to an explanatorily potent property that explains various instances of the phenomenon; in the latter, some feature that is present in each case, but does *not* explain the phenomenon, is exploited to formulate a generalisation. It is the deflationist's contention that the truth predicate only plays the latter, merely generalising, role in the claim that true belief facilitates success.

Mere generalisation is an odd tool. Most of the time, if we generalise over particular instances of a phenomenon by picking up on an explanatorily *irrelevant* feature, what we'll end up with is just a bad generalisation. I might, for example, generalise over Alison's and Callum's practical success by exploiting the fact that both have 6-letter names: 'having a 6-letter name facilitates practical success', I might say. This is a terrible generalisation; and it is so precisely *because* the feature so highlighted is explanatorily impotent when it comes to practical success. Good generalisations

are typically a sign that the property in question *is* explanatorily relevant; and the truth generalisation is a good one.

There are, however, cases where good – or, at least, better – generalisations can be formulated using explanatorily irrelevant features, since there can be *correlation* without explanation. One way in which this is possible is if the correlation is merely coincidental. For example, a disproportionate number of recent U.S. presidents have been left-handed. We might formulate a generalisation thus: ‘left-handedness facilitates electoral success in the U.S.’; and insofar as we are able to read this as a *mere* generalisation, it might be true enough.<sup>16</sup> But mere generalisation in this case is only permissible because there is no explanation forthcoming, i.e., because the correlation is coincidental. Presumably, then, the deflationist will want to deny the analogy between being left-handed and electoral success on the one hand, and true belief and practical success on the other; otherwise the correlation becomes brute and inexplicable because coincidental. There are many coincidental correlations in this world, but this is not one of them.

However, the only kind of case I can think of in which good generalisations can be formulated without explanation, but where the relevant correlation is nonetheless plausibly non-coincidental, is when there is a third-factor explanation. For example, perhaps telesales executives who stand during their calls make more sales because the kind of person with the energy and drive to stand all day is also more likely to make more sales. We can offer the generalisation: ‘standing while making telesales calls facilitates successful sales’; and this might be a good (if potentially misleading) generalisation despite the lack of explanation. What *explains* the greater sales is the person’s drive; and since this also *explains* why they stand during calls, it plausibly explains the correlation.<sup>17</sup> As far as I can tell, then, if the deflationist is to maintain that truth does not explain practical success despite maintaining that this generalisation picks out a genuine and non-coincidental correlation, she must be committed to there being some such third-factor explanation.

This suggestion is quite implausible. We are to believe that there is some third factor, X, which explains why rational agents typically perform successful actions (presumably within certain parameters); and *also* explains why rational agents tend to believe the truth (presumably within those same parameters).<sup>18</sup> That is, we are to believe that we *could explain* the practical success of actions performed by rational agents entirely in terms of X; but instead (presumably for expressive ease) offer a generalisation formulated in terms of the merely correlated property of truth. Given belief’s role in the explanation of rational behaviour, the explanation of practical success in terms of X must go via the agents’ beliefs: X will explain why agents tend to have success-conducive beliefs within

---

<sup>16</sup> To this end, ‘facilitation’ is probably the wrong word, but given that the deflationist typically takes herself to be entitled to it in the truth case, it is unclear why we should not be entitled to it here. This, I think, just highlights how odd the deflationist’s suggestion is; and how this oddness is often masked by the language used. Note, below, that it would seem false to say that ‘standing up during calls *facilitates* telesales success’.

<sup>17</sup> I am granting this for the sake of argument; we ought to note the structural similarity here with the conjunctive quasi-explanations of one-off coincidences.

<sup>18</sup> X might most plausibly be a psychological (especially a belief-forming) mechanism. Wrenn (2011: 468) seems to have something like this in mind.

the relevant parameters. But despite the fact that X *also* explains why these very beliefs tend to be true, the explanation of the agent's success will *not* mention the truth of said beliefs. The truth of the beliefs thus becomes bizarrely epiphenomenal. And we might reasonably ask why it is that X, which produces success-conducive beliefs, also happens to produce true beliefs. Is it a coincidence? Why does it not produce success-conducive false beliefs? Are there none? Why not?

Either way, the main problem with this proposal is that it's simply implausible that there is any such X that can, alone (i.e., without reference to truth), explain why the agent tends to be successful. As mentioned, this explanation will go via the success-conducive beliefs, but not their truth (which is epiphenomenal). It must, rather, go via their *causal profile*; the beliefs will have the same causal profile whether or not they are true, after all. But the causal profile of the beliefs X produces can only plausibly explain *why the agent performs the actions she performs*; and, as emphasised, this is merely the first step in the explanation; it is not to explain *why the actions she performs are successful*. Explaining the *success* of the actions thus performed requires explaining why the actions tend to go on to satisfy her desires. There is nothing in the causal profile of the beliefs *alone* that explains this: it requires the contribution of the world.

Wrenn (2011: 468-9) is surely right when he points out that evolution will select against those creatures whose beliefs tend to lead to unsuccessful actions. Presumably, then, evolution might select for some X that produces success-conducive beliefs; and this might “explain” why X produces success-conducive beliefs. But this evolutionary “explanation” is phenotypical, not genotypical: we are told what feature the beliefs and production mechanisms are selected for (producing successful action), but not what property they possess that gives rise to this feature. Why is it *these* beliefs, selected for by *this* mechanism, that lead to successful action? (Granted a cheetah's legs are selected for running fast; but what is it about *these* legs that enable it to run fast?)<sup>19</sup> The obvious explanation, given the correlation with true beliefs, is that X is selected for producing *true* beliefs precisely because *true* beliefs are useful for bringing about practical success. It is not merely the beliefs' causal profile that explains the practical success; but the beliefs standing in the right relation to the world. But then X is not a third-factor after all: X explains why the agents tend to have true beliefs, which in turn explains why the agents tend to be successful. This picture is far more appealing, and once again it is only antecedent deflationary commitments that could push us to resist it.

There are other reasons to be sceptical of such third-factor proposals. For example, it is not clear that any third-factor explanation could have the robust modal profile that the truth generalisation possesses. Indeed, it's not even obvious that correlations with third-factor explanations are truly non-coincidental, or non-coincidental in the right way. But in any case, it is

---

<sup>19</sup> There are parallels – here and throughout – with the dispute in the scientific anti-/realism literature as to whether or not we need to postulate the truth of our best theories to explain their success. The critical difference is that here the truth of the relevant representations, the beliefs, is not in doubt. My present concern is to press the implausibility of *simultaneously* conceding the truth of the success-conducive beliefs we have while *denying* the explanatory relevance of said truth to that success.

not clear how a third-factor explanation could hope to solve the central problem of the paper: the Alison/Callum problem. We can stipulate that Alison and Callum are physically and psychologically identical, so if X explains why Alison has the (success-conducive) beliefs she has, then it explains why Callum has the (success-conducive) beliefs he has. Assuming that we are within the usual parameters, Alison's beliefs are true as usual; while, unusually, Callum's belief are false. But this is of no explanatory significance, as far as the success is concerned: X *alone* explains the practical success of any action (as well as, incidentally, the truth of Alison's beliefs). So, if X explains Alison's success, then it explains Callum's; and if it doesn't explain Callum's, it doesn't explain Alison's. In this way, a third-factor explanation cannot hope to break the explanatory tie between the coincidental and non-coincidental cases presented.

### 3.4 Conclusion

Maintaining that the sole role of truth in the claim that true belief facilitates practical success is as a device for generalisation requires showing that individual instances of practical success can be explained without invoking truth (or, at least, anything more than a relevant instance of the schema). The deflationist has traditionally offered seductive individualised explanations that seem to leave no explanatory work for truth. I have argued that these individualised explanations are problematically incomplete, as shown by the fact that we can give such individualised explanations when the practical success is merely coincidental. The contrast between a non-coincidence and a coincidence is an explanatory one; and I have argued that the deflationist's refusal to explain the non-coincidental instances of practical success in terms of the truth of the agents' beliefs renders her incapable of establishing an explanatory contrast between the cases. The inflationist, meanwhile, is entitled to the obvious explanatory resources required to recover the distinction. Pending some future deflationary response, this is a major dialectical victory for the inflationist. I think the deeper lesson here is that the deflationist's response strategy manifests an underlying confusion with regards to *explanation*. As we have seen, it is patently not enough to explain a *non-coincidental* phenomenon to explain each individual instance of it and then generalise over these explanations by highlighting some feature that they share in common, *but which is explanatorily irrelevant*. *Mere generalisation* is only appropriate for cases of *mere correlation* of one form or another. I'm not sure it has been appreciated that endorsing this "merely generalising" view hereby brings with it substantial – and, I have suggested, quite implausible – commitments as to the explanatory status *of the correlation* between true belief and practical success.

The deflationary construal of the principle that true belief facilitates practical success is thus to be rejected; and with it a global deflationary conception of truth. This establishes an explanatory role for truth that must be played by an inflationary truth property. But deflationism is still a live option for those truths outside the scope of the argument. In the next chapter, I argue first that this

property must be a realist truth property, but second that the Success Argument is limited in its scope. In particular, I show that it does not extend to moral truths, expressivistically-construed.

# CHAPTER 4

## EXPRESSIVISM AND TRUTH

---

‘Quasi-realism is not realism. The truth is not out there.’  
Lenman (2014: 243).

The primary purpose of this chapter is to set the stage for Chapter 5, where I argue that metaethical expressivists ought to endorse an epistemically constrained theory of moral truth. In the last chapter, I presented an anti-deflationary argument for the explanatory potency of truth. In Section 1, I argue that the Success Argument motivates a realist theory of truth in particular and that it has limited scope. In Section 2, I argue that moral truths, expressivistically-conceived, fall outside the scope of the Success Argument; nonetheless, in Section 3, I point out that in the moral domain truth still plays the *normative* role of determining which moral beliefs we ought to have. Chapter 5 then details an argument that an anti-realist truth property plays this role.

The primary point I need to make here as far as expressivism and truth are concerned is that moral truths fall outside the scope of the Success Argument. But this also affords an opportunity to (re)consider the relationship between expressivism, as a theory of moral thought and talk, and moral truth. The literature has pivoted from one extreme to another on this point: from proto-expressivists denying the very truth-aptness of moral discourse, to deflationists suggesting that their theory offers an easy “fast-track” solution to the principle objection to expressivism: the Frege-Geach Problem. I approach these issues in an exploratory spirit, and sketch a novel view: that *inflationism* about moral truth might offer a novel “fast-track” solution to the Frege-Geach Problem, if the expressivist can earn the right to it. My goal is to suggest a significant theoretical advantage that may come with endorsing the kind of anti-realist inflationism about moral truth I motivate in the next chapter. The reader shouldn’t worry if she’s unconvinced – the argument of Chapter 5 stands on its own terms – but the metasemantic program sketched here casts its conclusion in an auspicious light.

### 4.1 The Scope of Success

#### 4.1.1 *Success and Correspondence*

In the last chapter, I argued that we need to postulate an inflationary truth property if we are to explain the practical success of certain actions performed by rational agents, and in particular

if we are to understand the non-/coincidental contrast that is apparent between instances of success where the agent has true beliefs and those in which she merely has the right false beliefs. The primary goal of that chapter was to argue against the deflationary contention that truth is explanatorily impotent. But the argument is not just an argument for inflationism about truth, but for a particular *kind* of truth property. This, I think, became apparent in §3.3.1, when we considered how the deflationist may try to recover the non-/coincidental contrast by extending her explanatory schema. The relevant instances of practical success are those in which an agent performs an action *psychologically pertinent* to an outcome, and the action itself is *effectively pertinent* with respect to that outcome. The psychological pertinence of the action – which action the agent thinks will result in the outcome – will be explained by the agent’s beliefs. Its effective pertinence, by contrast, will typically be explained by how things actually stand, by “worldly” facts.<sup>1</sup> In the relevant non-coincidental cases, there will be a “matching” between those beliefs that explain the psychological pertinence and those worldly facts that explain its effective pertinence; this matching is what renders the action non-coincidental, so it is this matching that we should appeal to in the explanation. This, of course, is the hallmark of realist, correspondence theories of truth (§1.2.2).

To see that no anti-realist truth property (§1.2.3) could play this role, consider a final variation on our thought experiment. Take our protagonist Bill. Suppose that Bill thinks he will get a beer if he nods his head, and he thinks this because he believes that he is at a bar where the bartender knows his drink preference and introverted nature, has line of sight of him, and thus will bring him a beer when he nods. And suppose further that these beliefs are true. So Bill nods his head at  $t_0$ , and as a consequence receives a beer at a later time,  $t_2$ , for exactly the reasons he thinks he will. This is so far just an ordinary, non-coincidental instance of practical success.

But now suppose that, unbeknownst to Bill, his head is *also* attached to a Wallace-and-Gromit-style beer-dispensing device, so that when he nods his head, the machine deploys a beer. This process doesn’t stop Bill getting a beer from the bartender – his beliefs on that front are true, after all. But the machine does, let’s say, work rather faster than the bartender, so Bill’s nodding directly results in Bill getting a beer from the device at a time  $t_1$ ,  $t_0 < t_1 < t_2$ .

So, we have only one action, resulting from only one set of true beliefs. But this action is “doubly” effectively pertinent: it results in the satisfaction of Bill’s desire in two different ways. We thus end up with two instances of success: one at  $t_1$  and one at  $t_2$ .<sup>2</sup> Call the former  $\text{success}_1$  and the latter  $\text{success}_2$ . Now, despite these instances of success resulting from the same action, which *a fortiori* results from the same set of beliefs,  $\text{success}_1$  is coincidental where  $\text{success}_2$  is not. The challenge is to make sense of this explanatory contrast.

---

<sup>1</sup> For ease here, I will talk about beliefs and facts; but this should not be read as requiring beliefs to be primary truthbearers (§1.1.2) nor a fact-based correspondence theory (§1.2.2). Nor should anything much be read into use of the word ‘worldly’ here.

<sup>2</sup> The difference in time here is just a heuristic.

The correspondence theorist will maintain that the beliefs that explain the psychological pertinence of the action correspond (stand in a substantive relation to) the very facts that explain the effective pertinence in success<sub>2</sub>. But they do not stand in such a relation to those that explain the effective pertinence in success<sub>1</sub>. So, the correspondence of the beliefs explains success<sub>2</sub>, but does not explain success<sub>1</sub>.

How might the anti-realist account for this contrast? I submit that she cannot. Suppose that the beliefs in question are true in whichever sense the anti-realist prefers – they might be superwarranted – and suppose she denies that there’s any explanatorily potent relation that holds between the beliefs and the facts that explain the effective pertinence in the second case (as the correspondence theorist maintains, and the deflationist denies). The difficulty is seeing why the beliefs being superwarranted should explain success<sub>2</sub>, but not success<sub>1</sub>. After all, in *both* cases the action results from beliefs that are superwarranted – they are the same beliefs.

Of course, the correspondence theorist too will say that both instances of success result from beliefs that correspond to *some* facts. But it is not correspondence to *any old* facts that the correspondence theorist appeals to explain practical success – and thus to make sense of the contrast between success<sub>1</sub> and success<sub>2</sub> – but correspondence to *the facts that explain the effective pertinence of the action*. This relation holds with regards to success<sub>2</sub> and not success<sub>1</sub>. The anti-realist, by contrast, can only appeal to the property of *being superwarranted*; and here there is no difference between success<sub>1</sub> and success<sub>2</sub>. If there is no substantive relation that holds between the beliefs that explain the psychological pertinence and the facts that explain the effective pertinence *in one case but not the other*, then there doesn’t seem to be any substantive difference between the cases.<sup>3</sup> It is precisely because the realist incorporates worldly relata in her analysis of the nature of truth that she can account for the difference here; and it is this that the anti-realist, by definition, does not do.

The anti-realist might, I suppose, concede that there *is* a substantive relation that holds between the beliefs and the facts that explains success<sub>2</sub>, which doesn’t obtain with regards to success<sub>1</sub>, but deny that this is what the truth of the beliefs consists in. This may or may not be a coherent position (it is, in effect, to concede the existence of the correspondence relation, while denying the correspondence theory of truth), but it is difficult to see what could motivate it. In any case, it is sufficient for my “engineering” purposes (§1.4) to have shown that correspondence plays the relevant explanatory role.<sup>4</sup>

---

<sup>3</sup> I take the anti-realist’s position here to be analogous to the deflationist’s in the last chapter. Neither has the resources to say that there’s anything explanatorily significant about *those* facts explaining the effective pertinence in the non-coincidental case, as opposed to the facts that do so in the coincidental case.

<sup>4</sup> Another option, I suppose, might be to maintain that the “matching” that holds between the beliefs and the facts *consists in* superwarrant, since if truth *is* superwarrant, then a belief’s being warranted allows us to disquote. But this strikes me as a confusion. The point of the Success Argument is that the “matching” here has to be a *substantive* (i.e., explanatorily potent, sparse) *relation* that holds between the beliefs and the facts. Superwarrant is not one of these.



### 4.1.2 *The Scope of the Success Argument*

So, we have an argument in favour of realist inflationism about truth. If we were labouring under a monistic presupposition, then our primary task from here on would be to illuminate the nature of correspondence.<sup>5</sup> But with the multiplicity of pluralistic options canvassed above (§1.3.2) on the table, we need to be sensitive to the *scope* of the Success Argument.

Two points need to be noted. First, the “worldly” facts with which the beliefs purportedly correspond need to be capable of explaining the effective pertinence of the agent’s action; the objects and properties with which the truthbearer is concerned must therefore be capable of entering into such explanations. So, it may or may not extend to truthbearers concerned with abstract objects like numbers, since if an object is causally impotent, it may not be able to feature in the right kinds of explanation. It is also unclear whether or not it extends to those truthbearers that concern objects and properties with a narrow cosmological role (§2.2.3); so, if mind-dependent objects and properties have only such a role, then perhaps it does not extend to the relevant truthbearers. I’ll leave open just how wide the scope of the argument is.

The second point is that the relevant *mental states* have to be capable of explaining the psychological pertinence of the action. On a familiar Humean view, beliefs and desires play distinct roles in psychological explanation; they have different “directions of fit”. Beliefs have mind-to-world direction of fit, their job is to represent the world as being thus-and-so, but beliefs by themselves are incapable of pushing you into action. Desires, by contrast, are primarily motivational, and thus have world-to-mind direction of fit: their job is to push you into acting. The states that feature in the relevant psychological explanations are beliefs in this substantive sense; the truthbearers for which the success argument runs are thus ones that are appropriately related to such beliefs. In the next section, I argue that, given metaethical expressivism, moral truths are not so related.

## 4.2 **Metaethical Expressivism, Moral Truth, and the Frege-Geach Problem**

### 4.2.1 *Metaethical Expressivism*

In what follows, I assume hermeneutical metaethical expressivism is the correct explanation of the nature of moral thought and talk. The view as I present it is intended to be in keeping with the “quasi-realist” views of Simon Blackburn (1993, 1998a) and Allan Gibbard (1990, 2003), and thus to enjoy similar motivations and problems. But I by no means intend to give a close exposition of their theories. The theory as I present it (or, more accurately, outline it) is my preferred take on expressivism, and should accordingly be read in the first instance prescriptively, rather than

---

<sup>5</sup> No mean feat. I have nothing substantive to add to this debate here. My working example will continue to be the kind of causal/teleological interpretation posited by Lynch (§1.2.2), but I will not lean on this at all.

expositionally. Note that the argument of Chapter 5, while keyed to a particular expressivist understanding of moral fallibility, is not dependent on the particular version of expressivism to which I am most sympathetic; it should be general to any version of expressivism that utilises a similar theory of fallibility.

Expressivism as I understand it is a view in the philosophy of mind and derivatively the philosophy of language in the moral domain. The constitutive claim is the psychological one: that moral judgements are conative or affective, or generally desire-like, states. For example, to judge that eating meat is wrong might be to (come to) *disapprove* of eating meat; while to judge that giving to charity is right is to *approve* of giving to charity. To say ‘eating meat is wrong’ is thus to *express disapproval* of eating meat; to say ‘giving to charity is right’ is to *express approval* of giving to charity.

A core motivation for expressivism, from which it in part derives its anti-realist credentials, concerns how it promises to explain the existence of moral thought and talk, which is without invoking a distinct class of moral properties that we have evolved to be representationally responsive to. This is best captured by Blackburn:

‘Animals with standing dispositions to cooperate (say) do better in terms of other needs like freedom from fleas or ability to survive failed hunting expeditions by begging meals from others. No right, duty, or value plays any explanatory role in this history. It is not as if the creature with a standing disposition to help those who have helped it does well *because* it is a virtue. Its being a virtue is irrelevant to evolutionary biology. There is no such naturalistically respectable explanation...

Now contrast the kind of evolution already sketched with any that might be offered for, say, our capacity to perceive spatial distance. Again, what matters here is action. But what we must be good at is acting according to the very feature perceived... It is because our visual mechanisms show us far-off things as far-off and near things as near that we work well using them. That is what they are *for*. We can sum up this contrast by saying that although the teleology of spatial perception is spatial, the teleology of ethical commitment is not ethical. The good of spatial perception is to be representative, but the good of ethical stances is not.’ (1993: 169)

‘Here on the one hand, we are with likes and dislikes, sources of joy and pain, set in a social world, with real but limited sympathies, and limited but real needs to cooperate with each other. Here on the other hand is our language of valuation. Is it intelligible that beings of the kind we sketch should develop language of the kind we have, with its inputs, inferential rules, and outputs as we have them? Do we have to invoke divine sparks, skyhooks, faculties of intuition, cognitive powers beyond anything given by the five senses and general intelligence, to close the gap? Do we have to invoke ‘response dependent’ truth conditions,

where we are “really” to be represented as talking about ourselves, and thereby ‘making the service greater than the God’? Or can we make the evolution intelligible just by thinking of what we had to do, and therefore did? My hope, following Hume, is that we can do the latter.’ (2010: 301)

Where the naturalistic explanation of the existence of our capacity to perceive spatial distance has to refer to spatial distance itself – i.e., that very part of reality that the relevant region of our thought is *prima facie* about – the existence of our moral thought is not, according to Blackburn, to be explained in terms of distinctively moral features of reality, which our moral thought exists to represent.<sup>6</sup> This is a direct rejection of the kind of ontological difference that other pluralists appeal to: it is *not* that our moral thought and talk concerns a distinctive region or aspect of reality. Rather, our moral thought does not exist because it represents some feature of reality *at all*; its import is to be explained in terms of the need for ‘*coordination* broadly conceived’ between agents (Gibbard 1990: 26).

Note that it is explicitly *not* the claim that what *makes* actions right and wrong is their contribution or otherwise to coordination and hence evolutionary success; what makes actions right or wrong is a “first-order” normative question, and expressivism qua *metaethical* theory is silent on first-order questions. It is vital that we keep this broad anti-representational motivation in mind when considering the nature of moral truth within the expressivist framework.

#### 4.2.2 The Frege-Geach Problem

Expressivism as a school of thought started with a Bacchanalian revel of which Ayer (1936: ch.6) was the life and soul, and we are all to some degree or another still trying to cope with the resulting hangover. In the heat of the moment, our expressivist forebears made all kinds of metaethical claims – that moral language isn’t truth-apt, that there are no moral facts or properties, that there is no moral belief or knowledge – and, in the cold light of day, expressivists have spent a long time trying to figure out just how much of this they really want to be committed to. The principle worry about such negative theses is that they look highly revisionary: the vegetarian who thinks that eating meat is wrong will also think that it is *true* that eating meat is wrong, that it is a *fact* that eating meat is wrong (that being wrong is a *property* that eating meat has), and that they *believe*, or perhaps even *know*, that eating meat is wrong. And, of course, it looks like moral statements can appear in valid inferences (i.e., inferences that necessarily preserve *truth*) as much as any other kind of statement can. If expressivism is supposed to be hermeneutical – an explanation of our actual moral thought and talk – it looks like it cannot afford to be so revisionary.

---

<sup>6</sup> This should already make us suspicious of the applicability of the Success Argument in this domain.

Note, however, that thus far I have made none of these negative claims: I have only said that moral judgements *are* desire-like states, and denied that their existence is to be explained by citing moral properties that they represent. As we will see below, this does not rule out that there are moral properties, nor even that moral judgements are (also) beliefs, in some sense.

Nonetheless, the psychological claim naturally gives rise to the worry that expressivists are not *entitled to* the relevant claims regarding truth, validity, facts, properties, belief, and knowledge. We are now in the ballpark of the large cluster of worries that have been discussed under the heading of “the Frege-Geach Problem”. The worry here is broadly that expressivism will be unable to explain some features of moral thought and talk, and hence fail as a hermeneutical project. Formulating a neutral characterisation of the worry is tricky, however. To motivate the idea that there is a worry *for the expressivist*, as opposed to a standing explanatory debt that any theorist needs to pay off, requires contrasting the expressivist’s explanatory resources to those of her “representationalist” opponent; i.e., those who think that moral judgements *are* representational in the way the expressivist denies. And it’s not like there’s convergence in opinion within this camp.

For ease of presentation, then, let’s consider a concrete example. David Lewis’s (1975) “head-first” metasemantics seeks to explain the semantic features of language by “pairing” sentences with mental states that they express. Roughly, Lewis’s theory is that, within in a particular community, a sentence ‘a is F’ expresses the belief that a if F just in case there are conventions of “truthfulness” and “trust” operating within that community, such that one only utters ‘a is F’ if one believes that a is F (truthfulness), and tends to come to believe that a is F when one hears someone utter ‘a is F’ (trust). The semantic features of the sentences are then to be explained in terms of the beliefs that they express. Recall that beliefs are typically explained as two-place relations holding between believers and propositions (however construed – for Lewis, as one might expect, as sets of possible worlds); the sentence then voices the proposition that is (determines) the content of the belief that it expresses (§1.1.2). The sentence is true just in case the proposition it voices is true (the familiar primalist story). Importantly, this gives us a straightforward way of explaining the meanings of *complex* sentences: ‘p & q’ voices the proposition <p & q> by virtue of expressing the belief that p & q. This also gives us a neat *compositional* story: ‘p & q’ is related to ‘p’ and ‘q’ as <p & q> is to <p> and <q>. Such a theory seems straightforwardly compatible with validity-, fact-, property-, belief-, and knowledge-talk within the relevant discourse. We’ll use this as our paradigmatic “representationalist” metasemantics.<sup>7</sup>

Contrast this with the expressivist’s starting point. The expressivist explains (MW) as expressing the state of *disapproving of eating meat*:

(MW)            Eating meat is wrong.

---

<sup>7</sup> Lewis’s metasemantics is step one of a larger “head-first” metarepresentational project, the second step of which requires grounding the representational contents of the states expressed.

It is vital that it expresses this state, as ‘Grass is green’ expresses the belief that grass is green, and does not *report* it, as ‘I believe that grass is green’ reports that the utterer has that belief (while it expresses the belief that they believe that grass is green). This is what stops expressivism from collapsing into a naïve subjectivism. The problem is seeing how we might explain the semantic features of such sentences on this basis. Disapproval may not even be a propositional attitude: on the face of it, if I disapprove of eating meat there may be no particular state of affairs that I disapprove of, but just the action generally. (I might disapprove of eating meat even while correctly believing that no one has ever or will ever eat meat, as I in fact disapprove of machete-massacres in space.) If so, then the content of the state expressed is just: eating meat. Even if it is a propositional attitude, the state expressed will have the *wrong* truth conditions; e.g., <people eat meat>, or some such. (“Disapproval”, of course, is just a stand-in; but this problem will result whatever desire-like state is expressed.)

Note that this does *not* immediately entail that the state does not (also) have the right truth conditions. But it shows that the expressivist is not entitled to assume that it does.<sup>8</sup>

To foreground the difficulty, consider how the expressivist is to explain the meanings of *complex* sentences, like ‘eating meat is not wrong’. For the representationalist, this expresses a belief with the complex content <eating meat is not wrong>. But it looks like the expressivist needs to identify some *other* mental state that it expresses. And if she is to explain meaning compositionally, it looks like this state will have to, somehow, be understood compositionally from disapproval of eating meat. But it cannot express that the speaker *does not* disapprove of eating meat, for one may do *that* without thinking that eating meat is not wrong (because one has suspended judgement, or simply never thought about the issue). But nor does it express disapproval of *not* eating meat, for that is to think that not eating meat is wrong; i.e., that eating meat is *obligatory*, as opposed to merely permissible.<sup>9</sup>

This is how I will understand the Frege-Geach Problem: as the metasemantic difficulty of explaining the semantic features of moral sentences, especially complex sentences, given that the atomic sentences express desire-like states that do not have the right truth conditions.

A “slow-track” solution attempts to solve the problem by pairing each sentence with a desire-like state that it expresses.<sup>10</sup> For example, Blackburn (1984: 192-5) proposed that ‘if eating

---

<sup>8</sup> I therefore do not think – as Schroeder (2008, 2010b) does – that the rejection of truth-conditional semantics is constitutive of expressivism. On the contrary, I think (with Ridge 2014) that the expressivist *ought* to endorse truth-conditional *semantics*. The problem arises because the expressivist does not appeal to truth conditions in the *metasemantics*: in explaining the semantic features of the atomic sentences. The discussion here is indebted to Ridge’s book.

<sup>9</sup> For background on the so-called “negation problem”, see Unwin (1999, 2001) and Schroeder (2008: ch.2).

<sup>10</sup> A different slow-track approach would be to build in more structure into mental states expressed. This is Schroeder’s (2008) strategy: on his proposal, moral predicates express the conative state of *being for*; ‘x is wrong’, for example, expresses *being for blaming for x-ing*. This extra structure allows the expressivist to pursue a similar strategy to the representationalist: ‘x is not wrong’ expresses *being for not blaming for x-ing*, for example.

meat is wrong, then eating eggs is wrong’ expresses disapproval of the state of *both* disapproving of eating meat and *not* disapproving of eating eggs.<sup>11</sup> This higher-order attitude towards the attitudes expressed by the component sentences is useful insofar as it explains why it is incoherent to believe that eating meat is wrong, believe that if eating meat is wrong then eating eggs is wrong, but not believe that eating eggs is wrong. But Mark van Roojen (1996) pointed out that this makes ‘if eating meat is wrong then eating eggs is wrong’ *mean the same as* ‘it is wrong to both disapprove of eating meat and not disapprove of eating eggs’; and this looks mistaken since there is no *logical* inconsistency between the latter and both ‘eating meat is wrong’ and ‘eating eggs is not wrong’.<sup>12</sup>

#### 4.2.3 ‘A Panoply of Dignities’<sup>13</sup>

In denying moral truth-aptness, properties, facts, belief, and knowledge, the proto-expressivist does not directly threaten to be revisionary about the meanings of moral predicates, since she still permits us to use sentences like (MW). The revision threatens to strike at our concepts of truth, property, fact, belief, and knowledge, by *inter alia* restricting their application: in at least paradigmatic cases, if we are willing to assert a declarative sentence, then we are willing to make the associated truth-, property-, fact-, belief-, and knowledge-claims.<sup>14</sup> To resist revision, expressivists have been pushed to suppose that these *concepts* form something of a minimalistic web.

To start, consider Wright’s (1992: 27-9) minimalistic theory of truth-aptness, which has been called *disciplined syntacticism*.<sup>15</sup> On this view, a sentence is truth-apt just in case it (i) permits of grammatical embeddings in constructions like negations, conditionals, propositional attitude ascriptions, and truth ascriptions (has the right *syntax*), and (ii) the use of said sentences is governed by agreed standards of warrant (has sufficient *discipline*). Since moral discourse manifestly passes this test, it is on this theory truth-apt.<sup>16</sup> It is then natural to pair this with a minimalistic account of

<sup>11</sup> Note that Blackburn later modified his view: see Blackburn (1993: 188-197).

<sup>12</sup> Compare, however, Baker & Woods (2015), who argue that logical properties are primarily properties of *sentences*, and in particular those sentences that use logical vocabulary.

<sup>13</sup> This phrase is from Blackburn (1993: 168).

<sup>14</sup> It thus threatens to *derivatively* strike at our concept of validity, and in turn the practice of moral inference. One might reject this point on the grounds that there are other declarative sentences – metaphors, fictional stories, or Austinian performatives – that are not properly evaluated in terms of their truth (Blackburn 1998b: 159-60). But the right response is Wright’s (1998a: 187-8): that these sentences are still truth-*apt* (and apt for the other properties of interest) – are still *capable* of being true or false – even though there are typically norms other than truth operating on their assertion.

<sup>15</sup> Initially by Jackson, Oppy, and Smith (1994).

<sup>16</sup> Ironically enough, Wright’s original presentation was designed to exclude expressivism as a theoretical option. Consequently, in the mid-90s, there was a short but intense debate over whether or not the minimalistic theory of truth-aptness is antithetical to expressivism: see Jackson (1994), Wright (1994), Smith (1994a, 1994b), Divers & Miller (1994, 1995), Horwich (1994), and Jackson, Oppy, & Smith (1994), as well as Blackburn (1998b) and Wright (1998a). Unfortunately, this was largely framed against the presupposition that the expressivist intends to *deny* the truth-aptness of moral discourse. I am closest to agreement with Horwich in this dispute: I side with Divers & Miller in thinking that there is a minimal concept of belief which describes no robust causal profile, but consequently side with Smith, Jackson, and Oppy in thinking that minimalism about truth-aptness is congenial to expressivism (but for the *opposite* reason: because it means they *can* allow that moral discourse is truth-apt).

<truth>,<sup>17</sup> at which point the expressivist has no incentive to deny that moral sentences are truth-apt, at least in the “everyday” sense of <truth> that is, on the minimal theory, voiced by our predicate ‘is true’.

We can then extend this minimalism to other nearby concepts. <Assertion> applies to the sincere utterance of any minimally truth-apt declarative sentence. The minimal theory of <fact> is much like the minimal theory of <truth>: to say ‘it is a fact that p’ is not to say much more nor less than one does by saying ‘p’. And since moral predicates will, on the expressivist’s theory, determine a function from possible worlds to extensions, they will ascribe properties, in at least the abundant sense. (On a minimal theory of <property>, the abundant sense is the one voiced by the singular term ‘property’.) Indeed, qua metaethical theory, as mentioned, the expressivist is neutral with regards to first-order normative ethics. So, it is quite possible that ‘is wrong’ even ascribes a comparatively sparse property. Suppose, for instance, that a naïve utilitarianism is the correct first-order theory, so for an action to be wrong is just for it to fail to maximise utility. The predicate ‘is wrong’ then ascribes the property of *failing to maximise utility*, and this can be a comparatively sparse property. Some such story can plausibly be given *whichever* first-order theory turns out to be correct.<sup>18</sup> And an action’s exemplifying this property can then constitute a fairly robust kind of *fact*, if there is indeed some such robust notion. (We might, for instance, cash out a robust notion of fact in terms of the sparseness of the properties involved.)

These latter points are significant when we consider the correspondence theory of truth. It is often assumed that the correspondence theory is incompatible with expressivism, I think due to the feeling that the correspondence theory is too “realist”. But it is difficult to locate any *in-principle* incompatibility between the theories that might back up this feeling of unease. An understandable assumption might be that, qua anti-realist, the expressivist cannot concede the relevant “worldly” facts and properties to stand in the correspondence relation. But the expressivist, we’ve just seen, will allow that there are facts and properties in *at least* a minimal, but perhaps also in a more robust, sense. So, there is no absence of truthmakers. Of course, the expressivist will want to deny that it is the job of moral *concepts* to be representational; moral concepts, after all, are to be cashed out in terms of their desire-like psychological role. So, there is probably an in-principle incompatibility between expressivism and the kind of causal or teleological interpretation of correspondence suggested by Lynch (§1.2.2): on these accounts, <wrong> refers to wrongness in virtue of, e.g., being tokened in the presence of wrongness, or having been selected for representing wrongness. But otherwise, I see no in-principle incompatibility between the theories.<sup>19</sup>

---

<sup>17</sup> §1.1.1 and §1.2.1. As stressed there, minimalistic or “deflationary” accounts of the *concept* of truth do not commit us to deflationism.

<sup>18</sup> See Gibbard (2006) for the fully general version of this argument.

<sup>19</sup> Note that, as far as my pluralist intentions are concerned, it is only good news for me if it turns out that there is some such incompatibility that I have failed to locate.

The expressivist also has good reason to extend her conceptual minimalism to <belief>.<sup>20</sup> This can be seen most clearly via the phenomenon of Moore-paradoxicality. Moore-paradoxical sentences are sentences of (something like) the form:

(MP)            p, but I don't believe that p,

where 'p', as ever, is a schematic marker to be replaced by a declarative sentence, for example:

(MP-A)            Ajax is a cat, but I don't believe that Ajax is a cat.

What is striking about such sentences is that they are typically perfectly capable of being true, since I am fallible, and yet they are “unassertable” (at least outside of highly unusual contexts). Sincerely uttering (MP-A) would display a kind of incoherence of my behalf, such that my audience would struggle to interpret what on earth it is that I am trying to say. The familiar explanation of this interpretive roadblock is that the former conjunct expresses the mental state that the latter conjunct reports that I am not in. The sincere utterance of such a sentence thus reveals a “fractured” state of mind; and, importantly, ordinary speakers are *aware* of this. We thus attribute a sensitivity to the fact that the state expressed by the former conjunct *matches* that which is the subject of “anti-report” in the latter conjunct (forgive the neologism) to speakers in order to explain the interpretive roadblock.

It is, then, striking that *moral* instances of (MP) result in Moore-paradoxicality:

(MP-B)            Eating meat is wrong, but I don't believe that eating meat is wrong.

This is just as jarring as (MP-A). For our explanation to work in this case, ordinary speakers need to be aware that the mental state that is the subject of the anti-report matches that expressed by the first conjunct; but the former is explicitly a *belief*. Thus we are pushed to accepting that moral sentences express beliefs.<sup>21</sup>

This may seem problematic, as we have introduced expressivism as the view that moral judgements are conative or affective states, and it is vital for its motivation that these states be

---

<sup>20</sup> Horgan & Timmons (2006) were relevant trailblazers here.

<sup>21</sup> Moore-paradoxicality within the expressivist framework receives extended treatment in Woods (2014), which I draw on here. Note that Woods formulates an anti-expressivist argument by harnessing the absence of Moore-paradoxicality when the anti-report is formulated using the expressivist's preferred desire-like state, e.g., 'Eating meat is wrong, but I do not disapprove of eating meat' is not as jarring as (MP-B). This argument is yet to be satisfactorily addressed in the literature, and here is not the place to tackle it properly. It is clear, however, that the expressivist will need to argue that the ordinary speaker's sensitivity to the state expressed by the former conjunct being a state of disapproval does not extend to a sensitivity to its being so *under that description*.



understood as significantly different from states whose function is to be explained representationally, of which beliefs, understood as states with a certain functional role and mind-to-world direction of fit, are the paradigm. However, there is a natural response, which is to first maintain that ‘belief’, in ordinary language, voices a minimal concept  $\langle \text{belief}_{\text{min}} \rangle$ , such that any state that is expressed by an assertion is a  $\text{belief}_{\text{min}}$ . Importantly, there is nothing more to being a  $\text{belief}_{\text{min}}$  than being the kind of state that is expressed by an assertion; it doesn’t require that the state plays a particular functional role, for instance. This explains the incoherence of (MP-B). We then say that a proper subset of  $\text{beliefs}_{\text{min}}$  are the conative or affective states that the expressivist maintains moral judgement consists in. A disjoint subset will then be beliefs in a *maximal* sense –  $\text{beliefs}_{\text{max}}$  – where this is to be understood in terms of playing a representational functional role. This allows the expressivist to concede that moral sentences express beliefs, while retaining the contrast that is a necessary prerequisite of her distinctive psychological claim.<sup>22</sup>

This view might seem at odds with a popular view in the philosophy of mind: that our ordinary concept of belief is the “folk psychological” one which gives such states a distinctive functional role that contrasts with that attributed to desire-like states. But I think there are a number of plausible options for the expressivist here, which we do not need to decide between for present purposes. She may argue, for instance, that the functional role is merely part of our ordinary *conception* of belief, so that our ordinary conception describes only a proper subset of those states picked out by the concept. Or, she might argue that ‘belief’ is systematically polysemous between  $\langle \text{belief}_{\text{min}} \rangle$  and  $\langle \text{belief}_{\text{max}} \rangle$ . Note that the ambiguity here is nearly perfectly analogous to the one that Kölbel postulates between  $\langle \text{truth}_d \rangle$  and  $\langle \text{truth}_s \rangle$  (§1.3.1): the maximal concept is a “beefed-up” version of the minimal one; so this is a close polysemy, rather than any immediately problematic radical ambiguity. (A more radical option – though not, I think, a hopeless one – would be to reject that there is a minimal concept of belief, and explain Moore-paradoxicality via an error theory: we are systematically mistaken about which sentences express beliefs. Or, in the other direction, we might endorse a minor revisionism, such that we ought to replace the maximal concept of belief that our predicate actually voices with a minimal one, which is more in line with how we use the term.) Allowing that we have moral beliefs, in this minimal sense, in turn allows that we can have moral knowledge, in a similar, minimal sense. (This is why the ‘non-cognitivism’ label is to be abjured.)

So, there is a sense in which the expressivist ought to allow that we do have moral beliefs. But recall that the Success Argument requires that the beliefs play a particular role in psychological explanation: the representational role, rather than the motivational role (§4.1.2). The expressivist, as I understand her, explicitly denies that moral judgements are beliefs in this maximal sense. This is

---

<sup>22</sup> Compare Wright’s idea (1998a: 190), which draws on Blackburn, that it may be platitudinous that declarative sentences express *commitments*, which may have either direction of fit: commitment plays the role for Wright that I am suggesting  $\text{belief}_{\text{min}}$  plays. My point is that we need to postulate  $\langle \text{belief}_{\text{min}} \rangle$  or  $\langle \text{commitment} \rangle$  as the concept voiced by ‘belief’ in at least some ordinary uses of the word if we are to explain moral Moore-paradoxicality.

not, necessarily, to reject the correspondence theory of truth in the moral domain; on the contrary, above I was at pains to stress that I do not see any in-principle incompatibility (or other “implausibility”). My point here is that, by expressivist lights, the moral beliefs fall outside the scope of the Success Argument; so, the Success Argument does not motivate the correspondence theory of truth in the moral domain.

This is sufficient to block the applicability of the Success Argument to moral truths, expressivistically-conceived. Might we also block it on the “worldly” front, by denying that moral properties can appear in the relevant kinds of explanation? It is sometimes thought that the expressivist cannot concede the explanatory potency of moral properties, but this cannot be quite right. For if it should turn out that moral rightness is, say, maximising utility, then the property of rightness will have whatever causal profile maximising utility has. Nonetheless, it is tempting to say something like: the expressivist cannot concede the explanatory potency of moral properties *qua moral properties*. (Recall Blackburn: ‘It is not as if the creature with a standing disposition to help those who have helped it does well *because* it is a virtue. *Its being a virtue* is irrelevant to evolutionary biology’ (second emphasis added). It is the “virtue-ness” that Blackburn doesn’t want to appeal to, at least in the evolutionary explanation.) Perhaps her anti-realist motivation proscribes that it is an action’s rightness, rather than its maximising utility, that is explanatorily potent. This requires some fine-grained understanding of explanatory potency; in particular, I am intrigued by the suggestion that moral properties *qua moral* can only have narrow cosmological role.<sup>23</sup> This may provide grounds for resisting the Success Argument for moral truths on this front, but I do not have space (and nor do I need) to develop the proposal here.

#### 4.2.4 A “Fast-Track” Solution?

The rampant minimalism described in the last subsection has something of a blessing-and-curse status in the literature. On the one hand, some theorists seem tempted to think that it allows for an easy, “fast-track” solution to the Frege-Geach Problem; on the other, others worry that it risks collapsing expressivism into realism.

First, why might postulating the minimalistic web of concepts solve the Frege-Geach Problem?<sup>24</sup> The thought seems to be that, since the expressivist doesn’t deny the truth-aptness of

---

<sup>23</sup> Here is Blackburn on this topic: ‘Might we not say that actions derived from tokens of sentences such as ‘That is the right thing to do’ are typically successful, when they are, because whatever was picked out *was* the right thing to do? My answer to this is complex. I admit that in some circumstances we can say this. But I deny that they form a pattern. If our interest is explanation, we will reject such explanation in favour of ones that look at the underlying circumstances.’ (2013b: 128) See also §7.1.2, fn.18. Intriguingly, at times Blackburn seems to see explanations involving success as a way of determining which states *are* representational (e.g., maximal beliefs): ‘...whether or not we see a stance as representational is tied in with our best accounts of explanation (in particular, the explanation of success of action based upon it).’ (2013b: 130)

<sup>24</sup> On this, see especially Horwich (1993) and Stoljar (1993). Blackburn continually expresses sympathy with the idea that deflationism about truth offers a “fast-track” solution to the Frege-Geach Problem – for an early and a late example, see (1993: ch.10) and (2015: xvi-xvii).

moral discourse, or even that it expresses beliefs or ascribes properties, the Frege-Geach Problem simply doesn't arise. Things are not so straightforward, however. The Frege-Geach Problem does not derive from the denial of truth-aptness *per se*; it derives from how we seek to explain the meanings (in the metasemantic sense outlined above) of moral sentences. Notice that the worry as I described it did not rely on denying even that moral sentences, or the mental states they express, have truth conditions: it relied on the fact that, when described as desire-like states, the states expressed would not have the right truth conditions. This is a core difference between representationalist and expressivist metasemantic stories: the representationalist appeals to truth conditions in the metasemantics, in a way that the expressivist does not.

The minimalistic web explains, first and foremost, *why we are willing to apply* terms like 'true', 'fact', and 'belief' within the discourse in question: it is because the discourse behaves in the right way. In particular, our minimalistic story began with a minimalistic account of truth-aptness: Wright's disciplined syntacticism. This, however, leaves the behaviour *itself* unexplained. Part of the relevant behaviour for a sentence to be truth-apt is that it permits of embeddings into logical constructions, for example; but explaining how moral sentences do so – explaining the meanings of complex moral sentences – is precisely what the Frege-Geach Problem challenges us to say.

This point is perhaps made most forcefully by James Dreier (1996: 42-4). Dreier asks us to imagine a language in which 'Hiyo' is used as a device for *accosting* people: one can say, 'Hiyo, Bob' to accost Bob. We then make this a "predicate" by stipulating that 'Bob is hiyo' just means what 'Hiyo, Bob' means. 'Bob is hiyo' thus has the necessary *syntax* (that of a declarative sentence) to appear in logical constructions. While this is necessary for truth-aptness, Dreier's example makes it clear that it's insufficient: 'If a dingo is near, then Bob is hiyo' (to use Dreier's example) is plainly unintelligible, even if we're given the meaning of 'A dingo is near' and the "standard inference rule story" about the meaning of the conditional. That *moral* sentences permit of this kind of embedding is thus something that needs to be explained; and providing such an explanation *just is* to solve the Frege-Geach Problem. *Merely* endorsing the minimalistic web thus provides no "fast-track" solution to the Frege-Geach Problem.

*Perhaps*, however, we can run something like the Lewisian representationalist story using *minimalist* beliefs. (There will still be a question, of course, for each moral sentence, about which *maximal* state it expresses.) A key worry about this proposal, however, is that minimal beliefs and their truth conditions are then playing a substantive explanatory role, in explaining the meanings of moral sentences. It is, as I have stressed, consistent with a minimal construal of the concepts that they ascribe substantive properties; but if these properties are explanatorily potent, then we cannot make do with a deflationary construal – we're owed an account of the nature of the properties involved. Now, we do have a maximal construal of the minimal beliefs in the moral domain: they are desire-like states, e.g., states of approval and disapproval. But then we land back in the Frege-Geach Problem: these states don't have the right truth conditions.

This is why many expressivists are occupied with providing a slow-track solution to the Frege-Geach Problem; and I have no in-principle objection to the development of such a solution. However, my suggestion is that we may be able to use an inflationary theory of moral truth – providing we are entitled to it – to develop a different kind of fast-track solution.

The general proposal is, in a certain sense, that we restrict the head-first strategy across the board to *atomic* sentences. We then also require an explanation of the meanings of the logical constants, for instance in terms of their inferential role.<sup>25</sup> That is, we look at the inferential connections between the states expressed by complex sentences and those expressed by their components. (There is thus still a metasemantic role for the states expressed by complex sentences here.) We then explain the meaning of a complex sentence, not in terms of a mental state with complex content that it expresses, but *directly* in terms of the truth conditions of the component sentences and the meanings of the logical constants.

This is straightforward for ordinary representational discourse. For example, we explain the meaning of ‘Grass is green’ in terms of the truth-conditional content of the belief that grass is green, and analogously the meaning of ‘Snow is white’ in terms of the truth-conditional content of the belief that snow is white; but having obtained the truth conditions of these atomic sentences in this way, we explain the meaning of ‘Grass is green and snow is white’ truth-functionally in terms of the meanings of ‘Grass is green’, ‘Snow is white’, and ‘and’. To be clear: it is still part of the view that the conjunction expresses the belief that grass is green and snow is white; i.e., a belief with the same content as the complex sentence. But the proposal rejects that the metasemantic explanation in this case goes *via* this mental state.

We can then pull the same trick for moral discourse. First, we explain the meanings of atomic sentences in terms of the desire-like states they express. We then earn the right to a substantive theory of moral truth; for instance, in the next chapter, I argue that moral truth consists in the state that the atomic sentence expresses having a property that I call *weak super-stability*. We can use this to obtain a substantive account of the truth conditions of the atomic sentences: a sentence S is true at a world w just in case the state that S actually expresses is weakly super-stable in w, for example.<sup>26</sup> We can then explain the meanings of the complex sentences *truth-conditionally*; not in terms of the truth conditions of the states that they express, but in terms of the truth conditions of their components, and the meanings of the logical constants.<sup>27</sup> Our metasemantics thus goes truth-conditional “above the atoms”.<sup>28</sup> The structural point is that an inflationary theory of moral

---

<sup>25</sup> Field (1977) is an historically prominent example.

<sup>26</sup> This is rough, but hopefully is sufficient for illustrative purposes.

<sup>27</sup> A similar (perhaps, in the end, equivalent?) proposal might be developed using the form-restricted inflationary pluralism about truth that I develop in §6.2.1. This gives us a substantive theory of truth for any complex truthbearer, which we might then use to obtain substantive truth conditions.

<sup>28</sup> In §5.1.4, I argue that first-order moral negations (i.e., negations of atomic moral sentences) are also apt for weak super-stability. So, the proposal in this framework involves going truth-conditional above the *literals*. This requires an independent solution to the negation problem.

truth might entitle the expressivist to an inflationary account of the truth conditions of *atomic* sentences, which can then be used to explain the meanings of complex sentences compositionally.

This looks perfectly consistent with the anti-representational motivation outlined above: expressivists, as I understand them, do not invoke truth conditions in the explanation of the meanings of *atomic* moral sentences; the truth conditions come later in the explanatory order. It is consistent with this that we utilise truth conditions in explaining the meanings of logically *complex* sentences: the expressivist’s account of atomic moral sentences says nothing about the meanings of complex sentences – indeed, that’s the *source* of the Frege-Geach Problem! This provides a different kind of “fast-track” solution.<sup>29</sup>

We still need to be told which mental states are expressed by complex sentences, of course. But now the pressure is off somewhat, as these states do not need to discharge the metasemantic task of explaining the meanings of the relevant sentences.<sup>30</sup> For instance, van Roojen’s criticism of Blackburn’s proposal from above may not count against it merely as a proposal as to which state is expressed by the conditional. Indeed, within the proposal sketched here, we could explain the meanings of the two sentences differently while maintaining that they express the same mental state.

If a proposal along these lines could be worked out in detail, this would be a major boon for inflationism about moral truth within the expressivist framework, which I will argue in the next chapter the expressivist ought to endorse anyway. Nonetheless, I will not lean on the proposal any more in this thesis. First, considerable progress has been made on developing a slow-track solution,<sup>31</sup>

---

<sup>29</sup> As an aside: to what extent can the expressivist make sense of moral *propositions*? Given her first-order neutrality, I have said that the expressivist ought to be happy with ‘is wrong’ being necessarily co-extensional with (and thus ascribing the same property as) some purely descriptive predicate like ‘fails to maximise utility’. One way of saying what is constitutive of expressivism is that the *concept* <wrong> plays a particular role in our cognitive economy, one that is to be cashed out in terms of a desire-like state. To think of something under the mode of presentation captured by <wrong> is, roughly, to disapprove of it (with suitable adjustments made for if we’re merely entertaining the hypothesis, as opposed to endorsing it). If propositions are composed of concepts, then I do not see why this concept should not be a part of a proposition; but unfortunately this is only because the idea is obscure enough that it is difficult to adjudicate. A more concrete suggestion in a similar vein comes from Michael Ridge (2014), who aims to make sense of moral propositions using a Soamesian framework: roughly, among those cognitive event types that are propositions, there are types of peculiarly normative predications, where irreducibly normative concepts (those that play a certain desire-like conceptual role) are predicated of things. (Ridge’s story is more complex, since he endorses an “ecumenical expressivism”, where normative sentences express pairs of states, one desire-like, one belief-like.) If propositions are Russellian, then the proposition voiced by ‘x is wrong’ will be the same as that voiced by ‘x fails to maximise utility’; but the expressivist I think can be happy with that as long as the difference in conceptual role between the two concepts voiced is suitably acknowledged; for instance, these two different concepts might play a different role in *determining* which Russellian proposition is the content. Similarly, as is implicit in the above talk of substantive truth conditions, it looks like propositions as sets of possible worlds is consistent with expressivism here. So, it’s not obvious that there’s anything in the nature of propositions, however they are to be construed, that precludes moral propositions within an expressivist network. Again, the challenge is not with compatibility, but with *earning the right* to acknowledge them. Ridge’s Soamesian framework might allow us to make use of moral propositions prior to moral truth or truth conditions. My suggestion is to start with moral truth, and earn the right to truth conditions (and, if needed, propositions) on the back of that.

<sup>30</sup> This may provide a new way of understanding Blackburn’s claim that ‘the fast track can benefit from some of security achieved on the slow, and the slow track can make use of some of the short cuts of the fast’ (1993: 186).

<sup>31</sup> See, e.g., Silk (2015).

and here is not the place to evaluate such progress, and nor can I hope to make the case that the proposal sketched here is preferable. Second, I could be accused of begging the question against the deflationist by assuming a substantive role for truth conditions that the deflationist may well want to reject. Third, and most important for my purposes, is that while this might motivate inflationism about moral truth, it's not obvious that, by itself, it can tell us anything about the *nature* of moral truth; i.e., that it gives us any reason to prefer a particular inflationary truth property in the moral domain.

As to the “Problem of Creeping Minimalism”<sup>32</sup> – the worry that the minimalistic construal of <truth>, <assertion>, <fact>, <property>, <belief> collapses the distinction between expressivism and ordinary cognitivist realism – I have always been somewhat confused by this challenge. After all, it's simply not the case that the expressivist and realist will say all and only the same things. The realist will deny that moral judgements are desire-like states, for example. But, in any case, we ought to note that the truth pluralist in particular has a ready way of retaining the distinction. For suppose that moral truthbearers are not apt for correspondence. (This can be so on either a partial inflationism or a domain-restricted inflationary pluralism.) Then *at the level of property* we can cash out representational notions of assertion, belief, knowledge, and so forth in terms of correspondence in particular. If moral truth consists in some anti-realist truth property, then we can likewise cash out anti-representational notions of assertion, belief, knowledge and so forth in terms of this property. Alternatively, if moral truth is deflationary, then the difference will be that moral assertion, belief, knowledge, and so on is *merely* deflationary.

### 4.3 Expressivism and the Normativity of Truth

#### 4.3.1 A Normative Role for Truth

I have argued that it is a consequence of expressivism that the explanatory potency of truth with respect to practical success does not extend to the moral domain. There is, however, a distinct role for truth here, since it plays the *normative* role of saying *which* moral beliefs we ought to have.

Indeed, making sense of this normative role cuts to the very heart of another one of the core worries about expressivism, which concerns the *objectivity* or *mind-independence* of morality within the expressivist framework. We've seen that the expressivist resists collapse into a naïve subjectivism by utilising the distinction between *expressing* and *reporting* a mental state. Critics have, nonetheless, continued to worry that expressivism engenders an implausible subjectivism or mind-dependence about morality.<sup>33</sup> Many of these worries can be mitigated by arguing that, within the expressivist

---

<sup>32</sup> So-called since Dreier (2004).

<sup>33</sup> For recent in-depth discussions (and rejections) of the worries, see Schroeder (2014) on objectivism/subjectivism, and Köhler (2014) on mind-in/dependence.

framework, claims regarding the subjectivity or mind-dependence of moral facts or properties are not metaethical claims, but are first-order normative claims, internal to moral discourse. To say that, for example, the fact that torturing children is wrong is subjective or mind-dependent is to say that what makes such actions wrong is that we (say) disapprove of them; which is a moral claim on which metaethical expressivism is silent, and a deeply implausible (and unpleasant) one at that.

Nonetheless, a residual worry is bound to remain. For grant that our moral judgements do not, in the first place, function to represent moral facts or properties. How, then, can we make sense of the idea that our moral judgements might *go wrong*? We seem to be able to make sense of this third-personally: to think that *you've* gone wrong is just for me to think that  $\neg p$  when you think that  $p$ ; and the expressivist will have an explanation of what's going on there. But we can get to the worry when we frame things *first-personally*: how can *I* (that is: each of us individually) make sense of the possibility that *I* might be wrong? After all, if my judgement that eating meat is wrong does not function to represent the world, then it can hardly *malfunction* and *misrepresent* the world either. This question, regarding first-personal fallibility, is internal to moral discourse: how can the expressivist *make sense* of an expression of moral fallibility like (MW-F)?

(MW-F) I think that eating meat is wrong, but that might not be true.

If the expressivist cannot make sense of first-personal fallibility, then each expressivist is either committed to their own moral infallibility, or the threat of a radical subjectivism or relativism comes crashing back in with a vengeance. Neither option looks great.

The following chapter takes this challenge as a starting point. Utilising the account of fallibility proposed by Blackburn (1998a), and building on work by Egan (2007) and Elstein (2013), I argue that there are limits on what sense the expressivist can make of her own fallibility: that is, I argue that the moral truths are epistemically constrained, by the expressivist's lights. The best explanation of this, I maintain, is that moral truth is epistemically constrained; that it consists in a property that I call *weak super-stability*.<sup>34</sup> Together with the argument from the previous chapter, this constitutes what I in Chapter 2 called a “direct” argument for domain-restricted anti-/realist truth pluralism (§2.3): an argument that truth for some truthbearers consists in correspondence, and that for others it consists in an epistemically constrained truth property. Unlike the other arguments

---

<sup>34</sup> Lynch (2013a) has also argued that expressivists ought to endorse an anti-realist truth property, but the argument there is unconvincing. First, correspondence doesn't get a look in, presumably because Lynch's expressivist doesn't think that moral judgements “represent reality”; but, as I argued above, it's difficult to find any in-principle incompatibility here. Lynch's only reason for rejecting deflationism is the problem of creeping minimalism; but this simply overlooks *partial inflationism* as an option. As pointed out above, the partial inflationist has an easy way of avoiding the worry. I, like Lynch, can be viewed in the next chapter as following up on Blackburn's (1984: 198) suggestion that ‘given attitudes, given constraints upon them, given a notion of improvement... we can construct a notion of truth’. Lynch (2013a: 390) says of this: ‘that means, of course, that we should reject... the idea that moral truths are true in the same way as non-moral judgements’. But it is difficult to see exactly how that is supposed to follow.

considered in Chapter 2, I at no point rely on the intuitive “plausibility” or “implausibility” of a theory of truth for a particular domain; my argument is thus not vulnerable to the objections I pressed against my fellow pluralists. (On the contrary, I have done considerable work above to suggest that correspondence in particular *is* in-principle open to the expressivist.)

#### 4.3.2 *The Normativity of Truth for Belief*

The normativity of truth with regards to belief has a literature unto itself, so it’s important to close with a few notes on how what I have to say bears on the issue. There is quite the range of views in the literature, from those who think that there is *no* peculiarly doxastic norm of truth – any obligation towards or goodness in believing the truth is derivative of other norms – to those who think that being normed by truth is *constitutive* of the very nature of belief – that what individuates belief as a mental state is that it is normed by truth.<sup>35</sup> My view might seem, at first blush, to be in tension with both extremes, since it is a consequence of what is argued in the next chapter that a property I call weak super-stability provides the normative standard for the assessment of moral belief, but *only* moral belief, not belief generally. In fact, I think that everything I say can be rendered compatible with the full range of views here.

First, at one extreme, David Papineau (2013) – in a paper explicitly titled “There are no norms of belief” – denies that there is any general alethic norm on belief, but is nonetheless happy to allow that there may, *in certain cases*, be an obligation of some variety to believe the truth. It is just that this obligation will be derivative of some other norm.<sup>36</sup> For instance, Papineau suggests that surgeons have a moral obligation to have true beliefs about how to perform surgery; but this is just derivative of the fact that having such beliefs will help them to competently perform surgery, and that they have a moral obligation to competently perform surgery. My own inclinations here are towards this end of the spectrum: it strikes me that very little *argument* has ever been provided for thinking that there is a peculiar alethic-doxastic norm. (It is often taken as a truism, based on a controversial normative reading of the claim a belief is *correct* just in case it is true.<sup>37</sup> Most of the intuitive data, I think, can be readily explained by two simple observations: (i) the generic practical utility of true belief; and (ii) that we normally *want* to believe the truth. The two together are sufficient to provide something of a generic injunction against false belief.)

---

<sup>35</sup> Note that the norm here, to be plausible at all, is “objective” rather than “subjective”: subjectively, one most plausibly ought to believe in accordance with one’s evidence, which may differ from person to person; but presumably this would be because (in some *objective* sense) one ultimately ought to believe the truth, and evidence is how you get there.

<sup>36</sup> For similar views, see Papineau (1999), Dretske (2001), Bykvist & Hattiangadi (2007, 2013), Glüer & Wikforss (2009, 2013).

<sup>37</sup> A salient exception here is the argument provided in Shah (2003) and Shah & Velleman (2005), which is that the norm explains why the deliberative question of whether or not to believe p “immediately gives way” to the question of whether or not p is true. The primary fault with this argument is that the proposed explanation is a terrible explanation; see, e.g., Steglich-Petersen (2006: 506-7). It is also sometimes said to explain why we cannot form beliefs at will (Williams 1973: 148); but this supposed explanation is obscure to me too.



Nonetheless, I take it to be very plausible that there is *some moral* obligation towards or goodness in believing the *moral* truth, of the kind that Papineau suggests in the surgeon case. Exactly how to cash this out is then a substantive question, on which we need not take a stand here. There is presumably no obligation to believe *every* moral truth. My preferred approach is what Ferrari (2016) calls an *axiological* interpretation, as opposed to *deontic*: for any moral  $p$ , it is (perhaps *prima facie*) good or valuable to believe that  $p$  (perhaps *rather than* not believing that  $p$ , or believing  $\neg p$ ) just in case  $p$  is true (perhaps providing one forms a belief concerning  $p$  at all). It is only a very weak normative claim of this nature that I intend to be endorsing when I say that truth norms moral belief.<sup>38</sup>

At the other end of the spectrum are those who think the norm of truth is constitutive of the very nature of belief. So-called “normativists” are opposed to those who think belief can be individuated by something like its causal profile or functional role; a mental state only counts as a belief if it is normed by truth. For instance, Velleman (2000: ch.11) suggests that in some cases *imagining that  $p$*  might have the same causal profile as *believing that  $p$* ; the only difference is that someone who imagines that  $p$  when in fact  $\neg p$  is not thereby doing anything wrong (indeed, this is often a prerequisite of a good bit of imagining), whereas someone who believes that  $p$  when in fact  $\neg p$  is thereby doing something wrong.<sup>39</sup> I have very little sympathy for this very extreme view, but I think my arguments are consistent with even this. The direct consequence of my alethic pluralism by normativist lights is that there are two different kinds of belief: those that are normed by correspondence, and those that are normed by weak super-stability. (Note that our pre-theoretical concepts <belief> and <truth> need not be sensitive to these distinctions if minimal.) But I have already argued that the expressivist needs to hold that there are two different kinds of (minimal) belief: those that are maximal beliefs, and those that are desire-like states. Obviously applying normativism within my expressivist framework means these two distinctions map each other perfectly.<sup>40</sup>

More plausible interpretations of the norm of truth interpret it evolutionarily: that beliefs have the evolutionary function of being true. We do often talk about functions in normative terms: my watch *ought* to tell the time correctly; hearts *should* pump blood around the body. But I am in agreement with Dretske (2001) that this normative terminology should not be taken too seriously.

---

<sup>38</sup> It is sometimes said that deflationism cannot make sense of truth playing any normative role; see, e.g., Wright (1992: ch.1) on the norm of truth on assertion. However, deflationists have a seductive response to this challenge which is analogous to the response given to the Success Argument, described in detail in Section 1 of the last chapter – see §3.1.3, fn.10. As far as I can see, the quasi-explanatory phenomenon of coincidence has no normative equivalent, so no analogous difficulty arises.

<sup>39</sup> This time-slice of Velleman only commits himself to the “teleological” view that believing is accepting a proposition *with the aim* or intention of doing so only if it is true; but he comes around to full normativism in Shah & Velleman (2005). For similar (yet in some cases importantly different) views, see Wedgwood (2002), Boghossian (2003), Gibbard (2005), Zangwill (2005), Engel (2013).

<sup>40</sup> My form-restricted pluralism, defended in §6.2, may result in an “infinite proliferation” of kinds of belief on this view. Even this, I think, might be defensible, since the “complex” kinds of belief will be in some sense grounded in the “simple” kinds. But I am not especially concerned to defend this, as I am not at all sympathetic with extreme normativism.

In any case, I have argued that correspondence explains success, and it is success that is adaptive, so if any beliefs are supposed to be true it is only those apt for correspondence. It is thus precisely the consequence of the limited scope of the Success Argument that this teleological interpretation cannot be extended to moral beliefs; which is in keeping with the anti-representational motivation behind expressivism too, of course. Again, we already admit that there are different kinds of belief, so it is no problem if these different kinds of belief should be importantly related to different properties.

## **Conclusion**

In this chapter, I have argued that the Success Argument motivates a realist correspondence theory of truth for certain truthbearers, but that this argument does not extend to moral beliefs expressivistically-conceived. However, there is a distinct normative role for truth to play here, and I have suggested that the inflationary theory of truth provided in the next chapter may have the substantial benefit of providing the basis for a novel fast-track solution to the Frege-Geach Problem.

---

## CHAPTER 5

# WHY EXPRESSIVISTS CAN ONLY GO SO FAR WRONG

---

‘To show that these fears have no intellectual justification means developing a concept of moral truth out of the materials to hand: seeing how, given attitudes, given constraints upon them, given a notion of improvement and of possible fault in any sensibility including our own, we can construct a notion of truth.’

Blackburn (1984: 198).

In this chapter, I argue that metaethical expressivists ought to endorse an epistemically constrained theory of moral truth, and develop a novel such theory. My argument is based on a particular understanding of moral fallibility within the expressivist framework. There is an intimate connection between fallibility and truth. To acknowledge my fallibility is to acknowledge that (i) there may be some  $p$  that I believe that is not true (that I might be *mistaken*), and (ii) that there may be some  $p$  that I do *not* believe that *is* true (that I might be *ignorant*). This chapter starts by asking what sense the expressivist can make of her own fallibility. I argue that Blackburn’s account entails that the moral truths are epistemically constrained; indeed, I establish necessary and sufficient conditions on moral truth by expressivist lights. This limitation is explained and predicted by an epistemically constrained theory of truth, where other theories leave the explanatory debt unpaid.

Sections 1 and 2 constitute an extended argument that the moral truths are epistemically constrained. Section 1 draws on a dispute between Egan (2007) and Blackburn (2009) to establish sufficient conditions for moral truth and moral untruth; Section 2 tightens the net to establish necessary and sufficient conditions. In particular, I argue that all and only the moral truths have the property that I call *weak super-stability*. In Section 3, I address Egan’s charge that the expressivist is committed to an implausible “smugness”. Section 4 postulates weak super-stability as a theory of the nature of moral truth.

### 5.1 Fallibility and (Super-)Stability

#### 5.1.1 Fallibility for Expressivists

There is a *prima facie* challenge for the expressivist in explaining the meaning of first-personal acknowledgements of moral fallibility, like (1):

- (1) I believe that eating meat is wrong, but that might not be true.

At this stage of the dialectic, we cannot explain the meaning of (1) in terms of its truth conditions – it must be done in terms of the psychological state expressed. The difficulty comes in saying what state is expressed by the second conjunct, with (1) coming, as it does, from the mouth of someone who *in fact* disapproves of eating meat.<sup>1</sup> It is not the expression of any uncertainty as to whether or not I really disapprove of eating meat – I might be quite certain of that, and in any case it's not *that* that I'm worried about. But nor is it mere acknowledgement of the possibility that I might change my mind. If I did that, then, by my present lights, I would have come to have a false belief, not a true one.

Could the expressivist perhaps explain its meaning in terms of its expressing my having a credence of less than one that eating meat is wrong? The answer is: not straightforwardly, and for reasons influentially laid out by Michael Smith (2002). Variable confidence in a moral judgement is, after all, a belief-like feature to which the expressivist is not immediately entitled, but must earn. Desires also come in degrees, but Smith argues that “strength” of desire more readily maps onto a distinct aspect of moral judgement which he calls ‘Importance’. Suppose that I am equally confident in two moral judgements: that it is wrong to lie to my co-workers, and that it is wrong to lie to my loved ones, for example. I might nevertheless be more motivated to act on the basis of the latter judgement than the former, because I take lying to a loved one to be *worse* than lying to a co-worker. A natural explanation for the expressivist is that in the latter case I have a stronger desire (e.g., greater disapproval) than in the former case. But then variation in desire-strength maps onto variation in judgements of import, rather than variation in confidence.<sup>2</sup>

Simon Blackburn (1998a: 318) has formulated a compelling response to this difficulty:

“The problem comes with thinking of myself (or of us or our tradition) that I may be mistaken. How can I make sense of my own fallibility? Well, there are a number of things that I admire: for instance, information, sensitivity, maturity, imagination, coherence. I know that other people show defects in these respects, and that these defects lead to bad opinions. But can I exempt myself from the same possibility? Of course not (that would be unpardonably smug). So I can think that perhaps some of my opinions are due to defects of information, sensitivity, maturity, imagination, and coherence. If I really set out to

---

<sup>1</sup> The relevant sentence appears embedded here, of course, which is non-ideal at this stage of the dialectic; but all the same points could be made by considering ‘eating meat might not be wrong’ as asserted by someone who in fact disapproves of eating meat.

<sup>2</sup> The flipside of this worry is that the account of fallibility considered “doubles-up” as an account of variable confidence. For reasons of space, I do not discuss this here. Ridge’s (2015) response to this problem relies on an antecedent entitlement to variable confidence. I discovered Ridge’s paper only after having written this chapter, so I have been unable to engage with it in any detail.

investigate whether this is true, I stand on one part of the (Neurath) boat and inspect the other parts.<sup>3</sup>

Blackburn's central contention is that 'I can have a commitment to *processes* of belief formation and adjustment that outweigh my commitment to *p*.' (2009: 202-3) When I acknowledge the possibility that a moral opinion of my own may be mistaken, I acknowledge the possibility that better execution of these processes of belief formation and adjustment – improvement to my state of information, sensitivity, maturity, imagination, or coherence – might lead me to change my mind.

Let us quasi-formalise this idea. Call my belief set 'B<sub>WG</sub>'.<sup>4</sup> Given that I believe that *p*, *p* ∈ B<sub>WG</sub>. I have epistemic standards that I endorse, such as information, sensitivity, and so on. Label this set of standards, 'S<sub>WG</sub>'. Different sets of beliefs (B1, B2, and so on) may be better or worse by the lights of S<sub>WG</sub>. When I worry that my belief that *p* might not be true, I am worried that there is some B > B<sub>WG</sub> according to S<sub>WG</sub> such that *p* ∉ B (where '>' means *better than*). That is, I express uncertainty as to whether or not there is some *improving change* (improving by the lights of the standards I endorse) that I might go through, such that I lose that belief.

Some clarifications. I take it that it is very plausible that ordinary moral agents endorse such standards, though this endorsement will not normally be articulable. Note that these epistemic standards may well include meta-standards, for weighing improvements by the lights of different standards against each other. We should also be careful to distinguish what is an improvement by my lights, i.e., according to S<sub>WG</sub>, from what I *take* to be an improvement by my lights. I might slip up, and think my views have gotten more coherent, when in fact they have gotten less so. I might think I am getting more informed, when I am being brainwashed. It is the former that features in the account of fallibility, not the latter.

The role that my standards for improvement play in this account is indispensable. First, there must be some constraint on which changes get to factor into our theory of fallibility. My beliefs might change for any number of reasons: I might get hit on the head, or get indoctrinated, or go through some bad reasoning. But these aren't the kinds of changes to my beliefs that I should be worried about when I worry about my *fallibility*. To be relevant to the truth or otherwise of my beliefs, the changes have to be *improvements*. But the expressivist cannot at this stage appeal to the idea of an improvement *simpliciter*, which is something that brings us closer to the truth or the facts of the matter. For to do so is to make antecedent use of the moral truth in the explanation of the meaning of expressions of moral fallibility. That is why it is *my* standards that appear in the account of *my* expressions of fallibility.

---

<sup>3</sup> See also Blackburn (1993: 20-2).

<sup>4</sup> Throughout, I use my own belief set as the stock example. This is to emphasise that the considerations of this paper apply *first-personally* for *any* expressivist. The reader is invited to read 'B<sub>WG</sub>' as being their own belief set.

## 5.1.2 Egan’s Challenge

However, Andy Egan (2007) has pointed out that there is a problem with utilising my standards in this way. For suppose that, for some  $p \in B_{WG}$ , there is no  $B > B_{WG}$  according to  $S_{WG}$  such that  $p \notin B$ . That is, there is no change to my belief set that I would count as an improvement such that I lose that belief. Egan calls such a belief, *stable*. On the present account of moral fallibility, it looks as though I cannot make sense of a belief of mine being stable and yet in error. I cannot know *a priori* that any of my beliefs are stable in this way; but I can know *a priori* that, if any of my beliefs are stable, then they cannot be in error.

On the one hand, this upshot might just look implausible: ‘it is obviously not unintelligible for someone to suppose that something may be one of his own most fundamental and stable commitments, but not at all likely to be true, and certainly not on *that* account.’<sup>5</sup> On the other, Egan contends that it has implausible asymmetric first- and third-personal implications.

Egan’s argument is as follows. Take two agents, X and Y. X has belief set  $B_X$  and Y has belief set  $B_Y$ ; X has standards  $S_X$  and Y has standards  $S_Y$ . X and Y disagree over some moral matter,  $p$ ; that is,  $p \in B_X$  while  $\neg p \in B_Y$ . Now, suppose that there is no  $B > B_X$  according to  $S_X$  such that  $p \notin B_X$ , so  $p$  is stable for X; but, simultaneously, there is no  $B > B_Y$  according to  $S_Y$  such that  $\neg p \notin B_Y$ , so  $\neg p$  is stable for Y. Egan thinks that it is plausible that there may be such a situation because there are two degrees of freedom: X and Y may have wildly different starting beliefs, or wildly different epistemic standards.<sup>6</sup> On the assumption that X and Y disagree, and where there is (moral) disagreement there is error,<sup>7</sup> at least one of X and Y must be wrong. Therefore, there can be stable but false moral beliefs. But I cannot make sense of the idea that a belief that is stable *by my lights* might be false. There is thus ‘a sort of moral error to which others are subject, but against which I have an *a priori* guarantee of immunity.’<sup>8</sup> Following Egan, we will call this the “Smugness Worry”.

The Smugness Worry will be subject to scrutiny in Section 3. But it is important to be clear from the outset exactly what the “smugness” attaches to. Note that my “immunity from error” here does not amount to an inability to go wrong as to whether or not  $p$ . Even if  $p$  is stable in the sense that no change *that I regard as an improvement* will dislodge it, I may still change my mind as to  $p$  for whatever reason (hitting my head, bad reasoning, forgetting); and if I do so, I could then be in the

---

<sup>5</sup> Blackburn (2009: 204).

<sup>6</sup> Egan calls such disagreement in stable moral beliefs, ‘fundamental moral disagreement’. This is potentially misleading. We have a common sense notion of fundamental moral disagreement, which seems to concern disagreement about particularly important or central moral questions. Such disagreement is distressingly common; but Egan’s notion seems to me far more demanding, and hence the phenomenon much rarer. This point crops up time and time again below, but see especially §5.3.3.

<sup>7</sup> The expressivist may be able to side-step the worry by endorsing some kind of relativism. Since consistency with absolutism is a desideratum, I’m going to implicitly assume absolutism for the most part, but the distinction is largely orthogonal to my argument. I will suggest a relativistic spin on my view at the end of §5.3.3.

<sup>8</sup> Egan (2007: 210).

wrong. Nor is it particularly a smugness regarding one of *my* beliefs. For suppose that you also believe that *p*: then I can no more make sense of the idea that your belief is false than I can my own belief being false – after all, we agree! So it is not really *my belief that p* that I am being smug about. Rather, the smugness attaches to *my standards and beliefs as a starting point*. I cannot make sense of a belief being stable *according to my standards* and yet in error, but I can make sense of a belief being stable *according to your standards* and yet in error. The smugness lies in the fact that *my standards cannot* fail to lead me away from error, given my starting beliefs, in a way that *your standards might* fail to lead you away from error, given yours.

### 5.1.3 Improving Improvement

To this extent, then, I seem to be assuming that my standards are the *right* standards.<sup>9</sup> Blackburn's (2009) response to Egan is thus focused on giving the expressivist a way to distinguish *actual* improvement from merely *perceived* improvement (i.e., an improvement by the lights of SWG).<sup>10</sup> For instance, he concedes that the *a priori* of the following principle would be “unpardonably smug”:

‘(M) If something is entrenched in my outlook, in such a way that nothing I could recognize as an improvement would undermine it, then it is true.’ (2009: 205)

But he maintains that the following principle may well be *a priori*:

‘(I) If something is entrenched in my outlook, in such a way that nothing that is an improvement would undermine it, then it is true.’ (2009: 206)

And, further, there seems to be no reason that we shouldn't generalise (I) to all agents, and thereby remove the first- and third-personal asymmetry. But if she is not to make antecedent use of the moral facts, how is the expressivist entitled to this distinction between something *I* endorse as an improvement and something that is an *actual* improvement?

On a close reading of Blackburn's paper, we in fact find two quite different suggestions. Pertinent to our discussion, the first attempts to explain actual improvement in terms of truth:

<sup>9</sup> This isn't *quite* what's going on, as it is really my standards *in conjunction with my starting beliefs* that I'm apparently arbitrarily privileging.

<sup>10</sup> Another strand in Blackburn's response is that Egan's challenge focuses on moral error itself, rather than judgements about moral error. Köhler (2015) argues that Egan's challenge recurs in this setting. I sidestep this debate by focusing on judgements throughout. While I'm sympathetic with Köhler's conclusion, I think there is much more to Blackburn's response than the aspect he discusses.

‘... if [any belief] were false, then an improvement is clearly on the cards, namely replacing it with the truth. Deflationism about truth is quite compatible with [the generalised version of (I)]. A deflationist will interpret it as a generalization corresponding to the schema ‘if  $p$  then an outlook which includes [the belief that]  $\neg p$  is capable of improvement’ and under any acceptable interpretation of improvement, this will be something to be asserted.’ (2009: 206)

This is a very different standard for improvement to those previously considered; and for reasons already mentioned, this option simply isn’t available to the expressivist. Take a relevant instance of the schema: “if *eating meat is permissible* then an outlook which includes the belief that *eating meat is impermissible* is capable of improvement”. Blackburn fails to realise that the antecedent of this conditional constitutes an instance of the original challenge: how is the expressivist to explain its meaning, given that it comes from the mouth of someone who *in fact* disapproves of eating meat? Egan’s challenge is that Blackburn’s theory does not, on the face of it, extend to beliefs that are stable by the lights of SWG. On the supposition, then, that my disapproval of eating meat is stable in this way, what sense can I make of the antecedent? Blackburn’s proposal thus presupposes a solution to precisely the problem at hand!

Indeed, there is a clear circularity to this proposal. The original challenge was to make sense of the gap between what I *think* is the moral truth and the moral truth; the new challenge is to make sense of the gap between what I *think* are the right epistemic standards and the right epistemic standards. Blackburn’s response to the first challenge invokes the epistemic standards I endorse. If his response to the second challenge in turn invokes the moral truth, then we are back where we started. To put the point another way: if the expressivist was entitled to the relevant notion of the moral truth as distinct from what I believe in terms of which she can understand the idea of an improvement *simpliciter*, such as in the deflationary manner Blackburn suggests, then there would be no problem of moral fallibility to begin with.

Fortunately, Blackburn provides an alternative means for understanding how what I *perceive* to be an improvement might come apart from what is *actually* an improvement:

‘...why should my own standards for improvement be themselves exempt from critical scrutiny?’

Of course, while I am deploying particular principles, I am not at the same time scrutinizing them. But they may come under the microscope at any time... A very small tincture of self-doubt would entail silence, on my own part, about whether everything that *is* an improvement is something I *regard* as an improvement, or conversely.’ (2009: 206)<sup>11</sup>

---

<sup>11</sup> Blackburn seems to see this response as a something of a first-personal way of reaching the same conclusion as before.



The thought is that the expressivist can say more or less the same thing about her standards for improvement that she says about her moral beliefs: I understand my fallibility in terms of instability under improvement, but where improvement for a particular moral belief is understood in terms of  $S_{WG}$ , improvement for a particular standard *in*  $S_{WG}$  is understood in terms of the *other* standards in  $S_{WG}$ . Quasi-formally,  $S_{WG} \in B_{WG}$ , and for any  $s_i \in S_{WG}$ , I understand the potential untruth of  $s_i$  as the potential for there being some  $B > B_{WG}$  according to  $S^*$ , such that  $s_i \notin B$  – where  $S^* = \{s_j \mid s_j \in S_{WG} \ \& \ s_j \neq s_i\}$ . In this way, I can understand the potential falsity of *any* of my standards. And, derivatively, the potential falsity of *all* of my standards: this must be understood in terms of the potential for *incremental* replacement of each standard in turn, rather than a wholesale replacement of the lot.<sup>12</sup> Hence the Neurath boat metaphor is particularly apt here. (*How* the expressivist makes sense of this sceptical scenario will be important later on.)

Notice, however, that this does not straightforwardly entitle the expressivist to talk of *the correct* standards for improvement. However, it does enable me to understand how my own standards might not be the correct standards, which in turn enables me to understand how a belief might be stable by the lights of  $S_{WG}$  and yet be untrue after all: even if there is no change that is an improvement according to  $S_{WG}$  that displaces  $p$  from my belief set, there may be some *improved* set of standards,  $S$ , relative to which  $p$  is not stable. And even if  $p$  is also stable relative to  $S$ , there may be some improved version of  $S$ ,  $S'$ , relative to which  $p$  is not stable. And so on. Thus I can make sense of a belief being stable relative to  $S_{WG}$  and yet in error after all. And once we can make sense of that possibility, Egan's Smugness Worry – at least as formulated – does not arise.

Indeed, in a sense this response was built into Blackburn's account of moral fallibility from the beginning. For a judgement concerning my own fallibility is really a conjunction of two judgements: (i) the (descriptive, non-normative) judgement that there may be some change that is an improvement according to  $S_{WG}$  such that  $p$  is displaced; conjoined with (ii) a *normative* judgement that such a change would be an *improvement*. As such, the second aspect should fall within the scope of Blackburn's account of fallibility. The above discussion makes it clear how it does so.

Up until now, we've spoken of orderings on belief sets relative to standards, e.g.,  $B1 > B2$  according  $S_{WG}$ . Moving forward, it will be useful to have a more general notion of a *self-endorsed improvement*. For any belief set  $B$  and set of standards  $S \in B$ , a self-endorsed improvement is a change from  $B$  to  $B'$  such that  $B' > B$  according to  $S$ , or a change from  $S$  to  $S'$  concerning some  $s_i \in S$  such that  $S' > S$  according to  $S^*$  (where  $S^* = \{s_j \mid s_j \in S_{WG} \ \& \ s_j \neq s_i\}$ ). That is, either a change to the belief set that is an improvement by the lights of the standards in that belief set, or a change to those standards that is an improvement by the lights of the other standards. For ease of presentation, we shall say that  $B2 > B1$  just in case there is some *series* of self-endorsed improvements from  $B1$  to  $B2$ .

---

<sup>12</sup> Or, at least, there is no novel block to making sense of this possibility *from expressivism*. It may be incoherent for some other reason; see, e.g., Egan & Elga (2005).

In these terms, Blackburn's account of moral fallibility is as follows. If  $p \in B_{WG}$ , then when I acknowledge the possibility that  $p$  might not be true, I acknowledge the possibility that there may be some  $B > B_{WG}$  such that  $p \notin B$ .

#### 5.1.4 Super-Stability and Inaccessibility

However, just as Blackburn's response to Egan's Smugness Worry *as formulated* was built into his account of moral fallibility, it should be clear that a version of Egan's stability consideration will recur given this more sophisticated understanding. And this has been shown by Daniel Elstein (2013: §3.1). For suppose that my belief that  $p$  is stable relative to  $S_{WG}$ . Not only that, but it is stable relative to any improved version of  $S_{WG}$ ,  $S$ . And it is stable relative to any improved version of this improvement,  $S'$ . And stable relative to any improved version of this improved improvement,  $S''$ . And so on. That is, there is *no* series of self-endorsed improvement that dislodges  $p$  from my belief set, no matter how I improve my standards. Following Elstein, we will call such a belief, *super-stable*.

This establishes a sufficient condition for moral truth by expressivist lights: if a belief of mine is super-stable, then I cannot *make sense* of the suggestion that it is untrue. As before, I cannot know *a priori* that any of my beliefs are super-stable (though I strongly suspect that many of them, surrounding the wrongness of torturing children and so forth, are), but it looks like I can know *a priori* the conditional that *if* any belief is super-stable, then it is true.

Such a limitation looks inevitable, given the role that improvement plays in Blackburn's expressivist theory of fallibility, and the fact that the expressivist cannot make use of an antecedent understanding of the moral truth in understanding a notion of improvement *simpliciter*.

Now, Egan and Blackburn both focus heavily on the aspect of fallibility that I above called the possibility of being mistaken, but recall that fallibility has two aspects: not just the possibility of believing that  $p$  for some  $p$  that is not true, but also *not* believing that  $p$  for some  $p$  that *is* true; not just the possibility of being mistaken, but the possibility of being ignorant.<sup>13</sup> The expressivist will make sense of this in pretty much the same way. Take some moral  $p$  that I do not believe. I make sense of the possibility that it might be true in terms of there being some  $B > B_{WG}$  such that  $p \in B_{WG}$ ; i.e., that I will *come to* believe that  $p$  through self-endorsed improvement.

Consider, then, the converse phenomenon to super-stability: *inaccessibility*. Suppose that there is no  $B > B_{WG}$  such that  $p \in B_{WG}$ ; there is *no* improving change I might go through, such that I come to believe that  $p$ . By reasoning parallel to the former, I cannot make sense of the possibility that  $p$  is true. We thus have a sufficient condition for *untruth*.

This second limitation, linking the inaccessibility of a belief (via self-endorsed improvement) to untruth, has an interesting consequence, which is that the expressivist needs to be able to make sense, in principle, of "factually defective" moral discourse of one kind or another. For suppose that

---

<sup>13</sup> E.g., 'I don't think that eating meat is wrong, but it might nevertheless be true.'

I have either suspended judgement as to whether or not  $p$  is the case, or have simply never considered the question. (For a concrete example, imagine a moral dilemma like Sophie’s Choice.) So I neither believe that  $p$  nor believe that  $\neg p$ ;  $p \notin B_{WG}$  and  $\neg p \notin B_{WG}$ . But now suppose that *both*  $p$  and  $\neg p$  are inaccessible for me; there is no improving change I might go through such that I come to a verdict as to  $p$ . I then cannot make sense of the suggestion that  $p$  is true; but nor can I make sense of the suggestion that  $\neg p$  is true. In principle, then, there may be some  $p$  such that the expressivist cannot make sense of the suggestion that either it or its negation is true.

There are a couple of natural suggestions: perhaps  $p$  falls into a truth-value gap; or the truth of  $p$  is indeterminate. Either way,  $p$  is “factually defective”. There may not *be* any such  $p$ ; but the expressivist’s theory of truth ought to be able to allow for this possibility in principle.

We thus have a sufficient condition for truth and a sufficient condition for untruth. Call these together *the Strong Super-Stability Limitation*. Already, then, the moral truths are epistemically constrained. On its own terms, this limitation constitutes an explanandum of which the expressivist owes us an explanation. Egan’s Smugness Worry, reformulated, will start with the claim that I can make sense of a belief being super-stable by someone else’s lights and yet in error, or inaccessible and yet true, and argue for an *implausible* asymmetry between myself and others on the back of this. This worry can be rejected, but I will delay discussion until Section 3, as the considerations in Section 2 will allow us to deal with it thoroughly. In the next section, I tighten the net, establishing further limitations on what sense the expressivist can make of her own fallibility.

## 5.2 Tightening the Net

### 5.2.1 *Anti-Scepticism*

We have established the following:

#### *Ineliminability-Truth*

If there is no  $B \geq B_{WG}$  such that  $p \in B$ , then I cannot make sense of the possibility that  $p$  is true.

#### *Inaccessibility-Untruth*

If there is no  $B \geq B_{WG}$  such that  $p \notin B$ , then I cannot make sense of the possibility that  $p$  is not true.

On their own terms, these amount to a peculiarly expressivist variety of *anti-scepticism*, being logically equivalent to the following:

*Truth-Accessibility*

If I can make sense of the possibility that  $p$  is true, then there is some  $B \geq B_{WG}$  such that  $p \in B$ .

*Untruth-Eliminability*

If I can make sense of the possibility that  $p$  is not true, then there is some  $B \geq B_{WG}$  such that  $p \notin B$ .

Given my starting point, any moral truth is accessible through self-endorsed improvement, and any untruth eliminable. Note that making this case has *not* relied on there being some “ideal” endpoint where I believe all the truths and don’t believe any untruths, which is accessible to me via self-endorsed improvement.

We saw above that the expressivist can make sense of the possibility that *any* of her beliefs is in error; and derivatively of the suggestion that *all* of her beliefs are in error (modulo fn.12). To be able to make sense of this possibility is to have a healthy kind of open-mindedness: it is to always be open to finding out that you are wrong. There is, however, something distinctively *optimistic* about how the expressivist makes sense of it: it is in terms of the possibility that she might incrementally self-correct until none of her beliefs are left standing. (Note: it is not that her beliefs *must* or *will eventually* self-correct; but that they *can*.) That is, to *make sense* of the possibility of global error, she uses the *anti-sceptical* idea that the truth is in principle accessible to her via self-endorsed improvement. As Blackburn (1996: 94) puts it in another context:

‘What is not guaranteed by this kind of thought is the intelligibility of a different, radical kind of global error: the possibility that the truth might be nowhere that we can get to from here... [T]he idea that moral truth may be entirely and totally hidden from even our best efforts at improvement is not guaranteed to be coherent by reflection on those efforts and their structure.’

When Blackburn mentions places “we can get to from here”, he is of course not speaking merely causally – we might get anywhere from anywhere, merely causally speaking. He is talking about where I might get to via (self-endorsed) improvement from my present belief state, and casting doubt on the coherence of the suggestion that the moral truth might be completely inaccessible via such a process. What we’ve seen is that this suggestion is ruled incoherent on the expressivist’s account of moral fallibility. The Strong Super-Stability Limitation *just is* this anti-sceptical thesis.<sup>14</sup>

---

<sup>14</sup> By Lenman’s (2014: 240) lights, Egan’s “smugness” is a vice of excess to which there is an antipodal vice of deficiency: ‘a kind of moral pusillanimity, a catastrophic lack of confidence in one’s own moral convictions and commitments.’ I like to think of the expressivistic anti-scepticism developed here as the virtue between Lenman’s vices.

## 5.2.2 Weak Super-Stability and Inaccessibility

In fact, the expressivist is committed to a stronger variety of anti-scepticism. In this subsection, I introduce the *prima facie* limitation; in the following two, I consider and reject ways of resisting it. The discussion gets rather involved, and the reader may find the pictorial representations in *Figure 2* useful.

Suppose first that I believe that  $p$ , but that there is some improving change that I might go through such that I no longer believe that  $p$  (*Fig. 2a*); for example, suppose that if I was more informed about the environmental impact of animal farming, I'd change my mind regarding vegetarianism. Now, by expressivist lights, the possibility I worry about when I worry that my belief is not true is, in fact, realised. Is it, then, *a priori* that if a belief of mine is *unstable* in this way, then it is untrue? This would be a major cost, since it would mean that *every* self-endorsed improvement is an improvement *simpliciter*, and that is surely an implausibly strong commitment. What's more, if I were to go through the relevant improving change, I would presumably *then* be able to make sense of the potential truth of my belief.<sup>15</sup> How, then, do I *presently* make sense of the potential truth of my belief, given that it is unstable? Or, quasi-formally: how do I make sense of the possibility that  $p$  is true, given that there is some  $B > B_{WG}$  such that  $p \notin B$ ?

The natural suggestion here is that I make sense of the potential truth of any such unstable  $p$  in terms of *further* self-endorsed improvement. That is, even if there is some  $B > B_{WG}$  such that  $p \notin B$ , there may be some  $B' > B$  such that  $p \in B'$  (*Fig. 2b*). There is a nice consistency to this suggestion, since by the expressivist's own lights, if I were to go through the relevant improving change, then this is how I would make sense of the possibility that I was in the wrong; so I do not have to *change* how I make sense of the possibility. Even if my belief that  $p$  is unstable, then, I can make sense of the possibility that  $p$  is true.

But the dialectic immediately iterates. For suppose that there is in fact some such  $B$  and some such  $B'$ . How, then, do I make sense of the possibility that  $p$  is not true? I cannot reinvoke  $B$ . Once again, *further* self-endorsed improvement plays this role: there may be some  $B'' > B'$ , such that  $p \notin B''$  (*Fig. 2c*). And even if there *is* some such  $B''$ , I can make sense of the possibility that  $p$  is true because there may be some  $B''' > B''$  such that  $p \in B'''$ . And so on.

This all seems part-and-parcel of Blackburn's account of fallibility. But this way of understanding things has an obvious limit. For suppose that there is some  $B > B_{WG}$  such that  $p$  is *super-stable* at  $B$  (*Fig. 2d*). It seems, then, that I cannot make sense of the possibility that  $p$  is untrue, for then there is *ex hypothesi* no *further* self-endorsed improvement through which I change my mind about  $p$ .

If a belief is super-stable at some belief set  $B' \geq B$  in this way, we will say that it is *weakly* super-stable at  $B$ . A belief is *strongly* super-stable at  $B$  just in case it is also super-stable at  $B$  itself.

---

<sup>15</sup> As long as it doesn't immediately become inaccessible; but more on that in a moment.

(So we are rechristening super-stability as *strong* super-stability, to keep it distinct from *weak* super-stability.)<sup>16</sup> So it seems that, on Blackburn’s proposal, the expressivist cannot make sense of the possibility that a proposition is weakly super-stable at her own belief set and yet untrue.

The same considerations run, *mutatis mutandis*, if a proposition becomes inaccessible at some  $B > B_{WG}$ ; and we can likewise distinguish *weak* inaccessibility in this sense from *strong* inaccessibility. Thus we can see that the expressivist cannot, on the face of it, make sense of proposition being *weakly* super-stable at  $B_{WG}$  and yet untrue; nor *weakly* inaccessible at  $B_{WG}$  and yet true. Call this *the Weak Super-Stability Limitation*.

### 5.2.3 Divergent Improvements

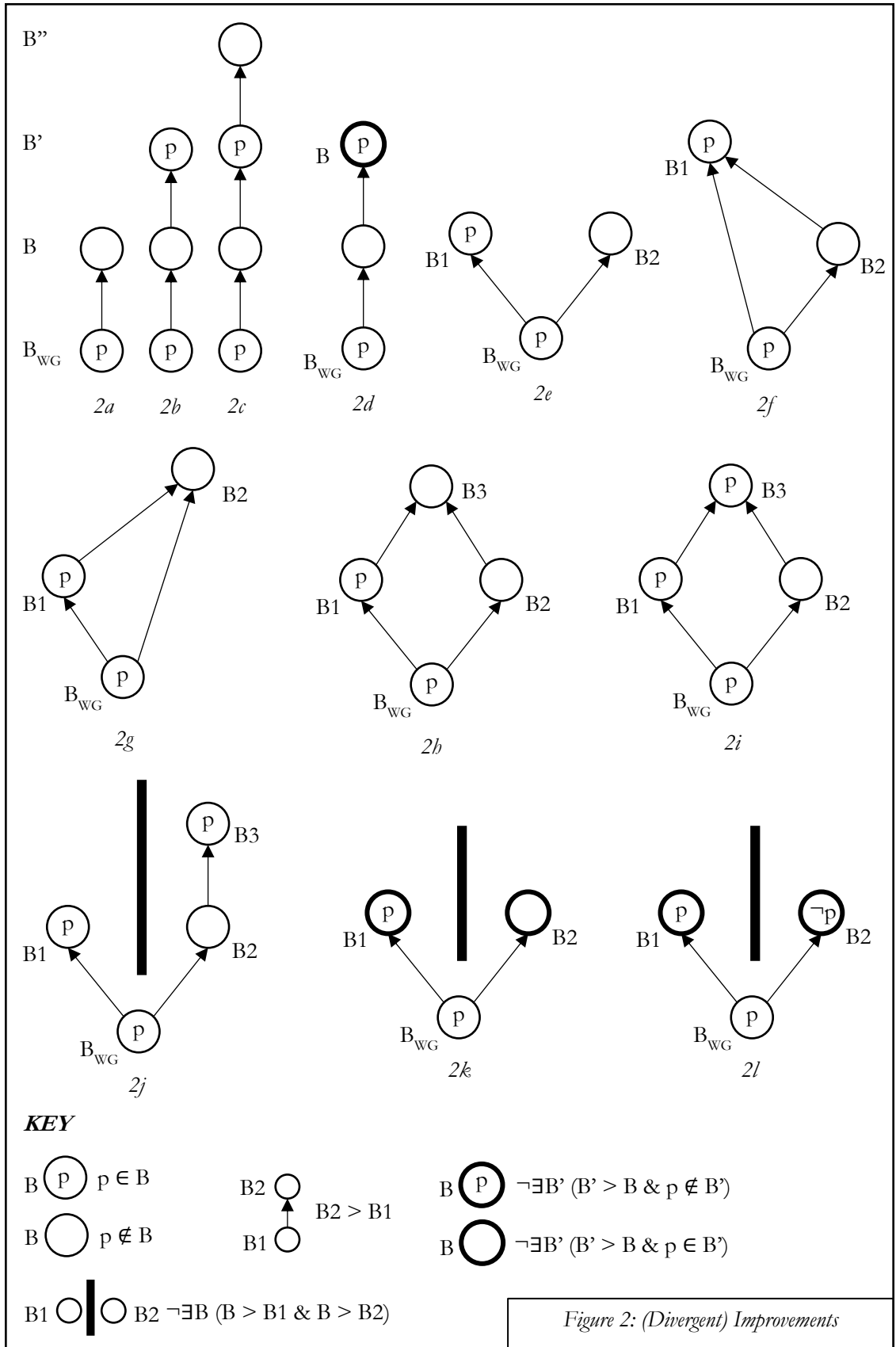
The natural way to resist the Weak Super-Stability Limitation is to try complicating the expressivist’s theory of fallibility in some way. A reasonable suggestion here is that I can make sense of the fallibility of those  $B > B_{WG}$  not just in terms of further self-endorsed improvement, but in terms of *divergent* series of self-endorsed improvement. There will be more than one way I can improve my beliefs. So, even if there is some  $B1 > B_{WG}$  such that  $p \in B1$ , there may be some other  $B2 > B_{WG}$  such that  $p \notin B2$  (Fig. 2e). And this is so even if  $p$  is strongly super-stable at  $B1$ . So, perhaps I can make sense of the possibility of  $p$  being untrue even if weakly super-stable in terms of a divergent series of self-endorsed improvement. Perhaps the existence of disagreement *among* those  $B > B_{WG}$  gives me a further way of understanding the fallibility of those  $B > B_{WG}$ .<sup>17</sup>

That there are divergent series of self-endorsed improvement from my starting point is guaranteed for me so long as there is more than one way in which I might improve my belief set. I might get more informed about the environmental impact of animal farming, and hence change my beliefs about vegetarianism; or I might instead become more informed about political inequalities, and hence change my mind about feminism. But the *mere* existence of some  $B2 > B_{WG}$  such that  $p \notin B2$  looks like a poor way to understand the fallibility of  $B1$  (given  $B1 > B_{WG}$ ,  $p \in B1$ ). Consider the divergent improvements just mentioned: the fact that I might not have changed my mind about vegetarianism if I had become more informed about something *irrelevant* looks, well, irrelevant. As long as there is more than one way to improve my belief set, there will always be such irrelevant divergent series of changes.

---

<sup>16</sup> In the rest of this chapter, I will sometimes talk in terms of *propositions*, rather than beliefs, being weakly and strongly super-stable. This is because the belief that  $p$  can technically be weakly super-stable for someone, even if they don’t believe that  $p$ ; while ‘weak super-stability’ is a term of art, this still sounds odd to my ear. But talking about the “proposition” here is just a way of talking about the belief, and nothing more should be read into it. (It’s also worth recalling that, as mentioned in §1.2.3, defining a truth property in terms of belief does not automatically render belief the primary truthbearer.)

<sup>17</sup> Compare Blackburn’s (1984: 197-202) discussion of a ‘tree... [where] each node (point at which there is branching) marks a place where equally admirable but diverging opinion is possible’ (199) and the resultant discussion of how to ‘*transcend the tree structure*’ (201). In some important ways, this chapter is a sophisticated and comprehensive discussion of the issues with which Blackburn is concerned there.



More important, however, is that B1 *itself* might be an improvement *on* B2 (Fig. 2f). More generally, while there may be divergent series of self-endorsed improvement from one and the same starting point, this divergence need not be permanent. Suppose, for instance, that I get more informed about animal farming *and then* get more informed about political inequalities; or vice versa. It is plausible that I might end up with the same beliefs either way. That is, even if B1 and B2 disagree, there may be some B3 such that  $B3 > B1$  and  $B3 > B2$ . Indeed, for small changes like these, it would be remarkable if there was no way of resolving the moral disagreement. We will say that B3 is a *point of convergence* for B1 and B2; and if two belief sets have a point of convergence, we will say they are *merely divergent*. If two belief sets are not merely divergent, they are *truly divergent*.

At the limit, the point of convergence for B1 and B2 may be one or other. Now, if  $B2 > B1$ , then B2 is a sensible way to understand the fallibility of B1; but this is just the “further self-endorsed improvement” understanding of fallibility that we are trying to find an alternative to (Fig. 2g). If  $B1 > B2$ , then it goes without saying that B2 is a poor way to understand the fallibility of B1, since B1 is an improvement on B2 (Fig. 2f again). Indeed, this is just how we understand the fallibility of B2!

Suppose, then, that there is some point of convergence B3 distinct from B1 and B2. Now, if B3 *disagrees* with B1 and agrees with B2, then this is a sensible way to understand the fallibility of B1; but, once again, this is just the “further self-endorsed improvement” understanding of fallibility that we are trying to find an alternative to (Fig. 2h). But if B3 *agrees* with B1 and disagrees with B2, then once again the existence of B2 is a poor way to understand the fallibility of B1, since B2 would come to agree with B1 through further self-endorsed improvement, which is just how we make sense of the fallibility of B2 (Fig. 2i). (There may, of course, be some divergent series of self-endorsed improvement from B2 to B4, such that B4 disagrees with B1 even if B3 doesn’t. But then the dialectic reiterates with regards to B1 and B4, instead of B1 and B2.) So, even if two divergent improvements on B<sub>WG</sub>, B1 and B2, disagree as to p, if there is some point of convergence between B1 and B2, then B2 cannot provide a plausible *alternative* way to understand the fallibility of B1; and thus no way to resist the Weak Super-Stability Limitation.

What this proposal would need, then, is for B1 and B2 to be *truly divergent*. But even this is not enough by itself. For even if  $p \in B1$  and  $p \notin B2$ , and B1 and B2 are truly divergent, there may nonetheless be some  $B3 > B2$  such that  $p \in B3$  (Fig. 2j); i.e., some improving change that B2 might go through such that it comes to agree with B1 with regards to p – and this is just how we understand the fallibility of B2. (Once again, there may also be some divergent  $B4 > B2$  such that  $p \notin B4$ , but then the dialectic reiterates with regards to B1 and B4.) So if B2 is to be a plausible way to make sense of the fallibility of B1, the disagreement must be more robust.

Suppose, then, that p is strongly *inaccessible* at B2 (Fig. 2k) or that  $\neg p$  is strongly super-stable at B2 (Fig. 2l), or both. Then, and only then, the divergent B2 perhaps provides a sensible way of making sense of the fallibility of B1.



What we've learned is that, if the expressivist is to make sense of the possibility of some  $p$  being weakly super-stable at  $B_{WG}$  and yet not true in terms of disagreement among those  $B > B_{WG}$ , then she must do so in terms of the possibility that  $p$  is *also* weakly inaccessible and/or  $\neg p$  weakly super-stable at  $B_{WG}$ . That is, there is one way of improving my beliefs such that  $p$  is strongly super-stable, but there is another way of improving my beliefs such that  $p$  is strongly inaccessible and/or  $\neg p$  strongly super-stable. True divergence between those  $B > B_{WG}$  is a necessary condition for this possibility.<sup>18</sup> There is some intuitive pull here, since if this situation obtains, then on what grounds could I rule that it is  $B1$  in the right and not  $B2$ ? (The same goes, *mutatis mutandis*, for making sense of the possibility that  $p$  is weakly inaccessible and yet true.)

#### 5.2.4 True Divergence as Scepticism

First, note that this only gets us so far. The expressivist still has no way of making sense of the possibility that  $p$  is weakly super-stable *and not weakly inaccessible (nor  $\neg p$  weakly super-stable)* and yet in error.

There are also multiple difficulties associated with trying to complicate the account of fallibility in this disjunctive way, i.e., making sense of the possibility that  $p$  is untrue when  $p \in B1$  for some  $B1 > B_{WG}$  in terms of the possibility that there is *either* some  $B1' > B1$  such that  $p \notin B1'$  *or* some  $B2 > B_{WG}$  such that  $p$  is strongly inaccessible at  $B2$ . Am I also to make sense of the fallibility of  $B_{WG}$  disjunctively? If not, then it seems that I should not make sense of my fallibility in a disjunctive way if were to go through the relevant change from  $B_{WG}$  to  $B1$ . So it seems that how I make sense of the fallibility of  $B1$  *changes* as I go through the relevant improvement, which looks untenable. On the other hand, if I am to make sense of my present fallibility in this disjunctive way, then what goes in the second disjunct? Is it just the existence of *any* belief set where  $p$  is strongly inaccessible? This looks too cheap (see §5.3.1); it looks akin to worrying that I might lose the relevant belief if I got hit on the head by a rock, or something otherwise quite irrelevant. So perhaps it is only the existence of belief sets *of a certain quality* where  $p$  is inaccessible that matters – but how do I identify these, if not in terms of the standards I endorse?

But most problematically, this way of making sense of the fallibility of  $B1$  *runs both ways*, and likewise undermines  $B2$ . As a result, the possibility of such true divergence through self-endorsed improvement amounts to a sceptical scenario. For suppose that  $p$  is strongly super-stable at  $B1$  and inaccessible at  $B2$ . If  $p$  is true, then  $B2$  is in a sceptical scenario that I cannot presently make sense of being in myself (given *Truth-Accessibility*); whereas if  $p$  is not true, then  $B1$  is in such a scenario (given *Untruth-Eliminability*). That is, this possibility guarantees that I can become *more* informed,

---

<sup>18</sup> In fact, formally speaking,  $p$  and  $\neg p$  might both be strongly super-stable at some  $B$  – we will have cause to mention this below, but it is largely irrelevant for our concerns here.

*more* sensitive, *more* mature, *more* imaginative, and/or *more* coherent, and *as a result* end up in a sceptical scenario that I cannot even make sense of being in at present.

Given the expressivist's anti-sceptical commitments regarding her own present epistemic situation, it is far from obvious that admitting this as a possibility is coherent. For instance, if I allow that a belief set which is an improvement by my own lights can be in such a sceptical scenario, then how can I rule out that I am presently in such a sceptical scenario? If anything, it seems that by my own lights I am *more* likely to be so, since I am epistemically less well-off! Furthermore, if I were to go through the relevant series of improvements to B1 or B2, I would then be unable to make sense of the possibility that I was in just the sceptical scenario that I would in fact be in. How, then, can I rule out that I am *presently* in such a scenario?

Indeed, it is not at all obvious what the expressivist *can say* about the truth of  $p$  in such a scenario; for each, one of the ways in which we make sense of their fallibility is realised, but then, intuitively, *both* should be in the wrong. If  $\neg p$  is strongly super-stable at B2, then we might make sense of this by supposing that neither  $p$  nor  $\neg p$  is true (that  $p$  is factually defective in some way). But if it is merely the case that  $p$  is inaccessible at B2, then what is the expressivist to say?

Not only, then, is trying to utilise this possibility in our theory of fallibility deeply problematic, but so is admitting its very coherence. Moving forward, then, I shall assume on the expressivist's behalf that such a scenario is incoherent, by their own lights.<sup>19</sup> This is to extend the expressivist's anti-sceptical commitments to the in-principle accessibility of any truth and the in-principle eliminability of any untruth to those  $B > B_{WG}$ , i.e., to any belief set accessible via self-endorsed improvement from  $B_{WG}$ . I do not think this is any great cost. On the contrary, if the expressivist cannot make sense of being in such a sceptical scenario at present, I do not see why she ought to be able to make sense of ending up in such a sceptical scenario by going through what she judges to be epistemic improvements. That is, given her *prior* anti-sceptical commitments regarding her own epistemic situation, this additional commitment does not render her view any less plausible; or so I have argued above. We want to allow that seeming-improvements might not be improvements *simpliciter*; but we can readily allow for that without conceding that they can lead us into a hopeless epistemic situation.

### 5.2.5 The Necessity and Sufficiency of Weak Super-Stability for Truth

If there is no true divergence among those  $B > B_{WG}$ , then the possibility for rival verdicts among those  $B > B_{WG}$  offers no new way of understanding fallibility. Barring any *other* way of making

---

<sup>19</sup> This would, of course, be problematic if true divergence were at all common; but see the discussion below (§5.3.3) concerning how low the bar for mere divergence is set. Indeed, I essentially leave full *defence* of this assumption to this later subsection, which I think bolsters my case beyond any reasonable doubt: we'll see that it is implausible that any but the most extreme views are truly divergent from my own; it is thus more implausible still that divergent improvements of my own are truly divergent from each other.

sense of fallibility within the expressivist framework, then, there is no way of resisting the Weak Super-Stability Limitation.<sup>20</sup>

Suppose, then, that  $p$  is neither weakly super-stable at  $B_{WG}$  nor weakly inaccessible. There is only one way in which this is possible: if, for every  $B \geq B_{WG}$  such that  $p \in B$ , there is some  $B' > B$  such that  $p \notin B'$ ; and for every  $B \geq B_{WG}$  such that  $p \notin B$ , there is some  $B' > B$  such that  $p \in B'$ ; i.e., I forever *change my mind* about  $p$  by improving my belief set. (Perhaps the easiest way to imagine such a scenario is via the idea of an “improvement loop”, i.e., a situation where  $B_2 > B_1$  and  $B_1 > B_2$ , such that  $p \in B_1$ , but  $p \notin B_2$ .<sup>21</sup> There must also be no divergent series of self-endorsed improvement where  $p$  becomes either strongly super-stable or inaccessible.) Let’s call such a proposition a *flip-flop* proposition.

I’m not entirely sure what the expressivist should say here. There certainly seems to be something problematic about *forever* flipping-and-flopping about a proposition through improvement. (Note that this isn’t *suspending judgement* about  $p$ , but *switching between* judging that  $p$  and not judging that  $p$ .) We might think that this shows that there is something wrong with the standards employed, in which case we’d expect improvement according to this meta-standard to eliminate this possibility; but that is *ex hypothesi* ruled out. Perhaps, then, the possibility of a flip-flop proposition is another sceptical scenario that the expressivist should rule out as incoherent.<sup>22</sup> It would be better to have something convincing to say on the expressivist’s behalf here; but for simplicity, I shall simply set aside flip-flop propositions as a tricky case for future investigation. Perhaps there is some way of spelling trouble for my arguments by making heavy weather of flip-flop propositions, but if so it is far from obvious to me how.<sup>23</sup>

Setting aside flip-flop propositions, every proposition must either be weakly super-stable at  $B_{WG}$  or weakly inaccessible at  $B_{WG}$ ; and, ruling out true divergence among those  $B > B_{WG}$ , cannot be both. Therefore, any proposition that is not weakly super-stable is weakly inaccessible, and vice versa. Since weak inaccessibility is sufficient for untruth, weak super-stability is thus also *necessary* for truth; and since weak super-stability is sufficient for truth, weak inaccessibility is also *necessary* for untruth. We have thus arrived at necessary and sufficient conditions for truth by the expressivist’s lights: weak super-stability at  $B_{WG}$ .<sup>24</sup> Call this the Weak Super-Stability Hypothesis.

<sup>20</sup> Since the Weak Super-Stability Limitation entails the Strong, it supersedes it for our purposes. But recall that we only need the Strong to show that the moral truths are epistemically constrained.

<sup>21</sup> Thanks to Gary Mullen for bringing improvement loops to my attention.

<sup>22</sup> Blackburn considers such a case at (1993: 22), and comes very close to endorsing the view I defend here.

<sup>23</sup> If one could develop an expressivist theory of truth itself (see e.g., Schroeder 2010a, Gibbard 2012, Ridge 2014), then an interesting example of a potential flip-flop proposition might be that expressed by the liar sentence, ‘This sentence is not true’. Perhaps this is a way of cashing the idea that ‘is true’ voices an inconsistent concept within an expressivist framework. While fascinating, I must leave investigation to a future occasion.

<sup>24</sup> If flip-flop propositions are untrue, then weak super-stability is *still* necessary and sufficient for moral truth, though weak inaccessibility will no longer be necessary for untruth. That is partly why I am not especially worried about flip-flop propositions for present purposes. So I will focus exclusively on weak super-stability, rather than weak inaccessibility from now on. (My worry is that ruling a flip-flop proposition untrue is arbitrary, given that for any  $B > B_{WG}$  such that  $p \notin B$ , there is some  $B' > B$  such that  $p \in B'$ .)

The Weak Super-Stability Hypothesis constitutes an explanandum. In Section 4, I will propose that the expressivist pay off this explanatory debt by understanding moral truth in terms of weak super-stability. First, however, we must confront Egan’s Smugness Worry.

### 5.3 “Smugness”

#### 5.3.1 *The Superficial Smugness Worry*

I have been arguing that the Strong and Weak Super-Stability Limitations *on their own terms* amount to a plausible enough variety of anti-scepticism. One may object to anti-scepticism, of course; but it is a tough dialectical job to argue that a position is implausible on the grounds that it *doesn’t* have sceptical consequences, especially within the expressivist’s anti-realist framework.

Egan’s Smugness Worry, however, is that these consequences arbitrarily privilege my own belief set over others’. After all, I have argued that a proposition’s being weakly super-stable *at B<sub>WG</sub>* is necessary and sufficient for its truth (by my lights); and that is at least formally consistent with distinct propositions being weakly super-stable at some other belief set and yet in error or not weakly super-stable and yet true. What makes my belief set so special?

The first thing to note is that *mere* smugness is not much of a charge. Blackburn has long distinguished what we might call the *superficial* problem of relativism from the *deep* problem of relativism. The superficial problem is

‘the vague and unfounded disquiet that I have no right to judge unfavourable people with any other opinion – those who practice human sacrifice, or murder Jews, for instance. Of course I have. My attitudes, and those involved in any system I could conceive of which might be superior to mine, alike condemn them.’ (1984: 199)

The point in the present setting is that the mere existence of *some* person, real or imaginary, who is subject to some error that I am not subject to need not bother me. Recall that all my “smugness” amounts to is thinking of someone else that there may be some moral truth they cannot access through self-endorsed improvement, or some untruth they cannot so eliminate. Those who “practice human sacrifice, or murder Jews” might have such pathological outlooks. Being “smug” with regards to such opposition in the sense mentioned is, I submit, unproblematic.

‘The deep problem [of relativism] is the suspicion that other, *equally admirable* sensibilities, over which I can claim no superiority of my own, lead to divergent judgements. This does take away my right to think of mine as true, which is equivalent to unsettling my commitments.’ (*ibid.*, emphasis added)

The task facing the would-be objector in this setting is to make the case that some *admirable* sensibility is liable to a kind of error to which I am immune; this would be a *deep* Smugness Worry. As we will see, the present framework allows us to get a much firmer grip on the problem, and I shall argue that those who would make this charge have some very hard work to do. We hereby shift the burden of proof firmly onto the objector.

### 5.3.2 Establishing a Network – Some Formal Results

First, some simple but important formal results show that divergence in which propositions are weakly super-stable at a belief set is inextricably intertwined with the possibility of true divergence:

*Theorem #1:* If B1 and B2 are merely divergent, then: if p is strongly super-stable at B1, p is at least weakly super-stable at B2.

*Theorem #2:* If p is strongly super-stable at B1 and  $\neg p$  is strongly super-stable at B2, then as long as B1 and B2 have reasonable standards, B1 and B2 are truly divergent.

*Theorem #3:* If p is weakly super-stable at B1 and  $\neg p$  is weakly super-stable at B2, then as long as B1 and B2 have reasonable standards, then there is some  $B1' \geq B1$  that is truly divergent from some  $B2' \geq B2$ .

The proofs are as follows:

*Proof #1:* Suppose that p is strongly super-stable at B1. Therefore, for any  $B1^* > B1$ , p is strongly super-stable at  $B1^*$ . If B1 and B2 are merely divergent, there is some point of convergence B3, such that  $B3 \geq B1$  and  $B3 \geq B2$ . Therefore, p is strongly super-stable at B3. Therefore, p is (at least) weakly super-stable at B2.

*Proof #2:* Suppose p is strongly super-stable at B1 and  $\neg p$  is strongly super-stable at B2. Therefore, for any  $B1^* > B1$ , p is strongly super-stable at  $B1^*$ ; and for any  $B2^* > B2$ ,  $\neg p$  is strongly super-stable at  $B2^*$ . Suppose for *reductio* that B1 and B2 are merely divergent, i.e., there is some point of convergence B3, such that  $B3 \geq B1$  and  $B3 \geq B2$ . Therefore, p is strongly super-stable at B3 and  $\neg p$  is strongly super-stable at B3. That is, B3's standards are such that there is *nothing* that B3 would count as an improvement that would eliminate this contradiction. Call such standards *unreasonable*; and call any standards that give rise to

unreasonable standards through self-endorsed improvement *unreasonable*. Therefore, if B1 and B2 have reasonable standards, there is no such B3. Therefore, B1 and B2 are truly divergent.

*Proof #3:* Suppose  $p$  is weakly super-stable at B1 and  $\neg p$  is weakly super-stable at B2. Therefore, there is some  $B1' > B1$  such that  $p$  is strongly super-stable at B1'; and there is some  $B2' > B2$  such that  $\neg p$  is strongly super-stable at B2'. By *Theorem #2*, B1' and B2' must be truly divergent if they have reasonable standards.

“Unreasonable” standards are those that allow both  $p$  and  $\neg p$  to become strongly super-stable. This can only happen if we abandon the standard of consistency; for otherwise there will always be an improvement on the cards that eliminates this contradiction. So unreasonable standards are those that either do not endorse consistency, or give rise (through self-endorsed improvement) to standards that do not endorse consistency.

What these results show is that any reasonable belief set, B, falls within a *network* of belief sets where *all and only the same propositions are weakly super-stable*, i.e., those that are merely divergent from B and its improvements, and that remain so through self-endorsed improvement.<sup>25</sup> This is a highly significant result. For there can be no question of smugness on my behalf over those within my network: there is no “sort of moral error” to which they are subject and I am not. (Indeed, my suggestion below is that the expressivist ought to understand moral truth as weak super-stability *within her network*, as opposed to weak super-stability *at B<sub>WG</sub>*.) Smugness only extends to those *outside* my network. So how smug the expressivist is being depends entirely on how extensive her network is. At the limit, if *every* belief set is part of her network, then there is no question of smugness.

### 5.3.3 Against True Divergence

The Deep Smugness Worry thus requires that there is some admirable belief set outside my network, which in turn requires that it is truly divergent from mine; i.e., that there is no way, even in principle, that we might improve our beliefs in accordance with the standards we endorse such that we come to agree.<sup>26</sup> The objector owes us some reason to think this is at all plausible. I hope that laying this bare renders plain the difficulty of the task. But it is worth running through in some detail.

<sup>25</sup> Merely formally speaking, if truly divergent belief sets are accessible via self-endorsed improvement from one and the same starting point, then a belief set may fall outside of its own network, on this definition. As I have suggested that the expressivist must rule out this possibility for herself on independent grounds, this doesn't much matter here. Thanks to Robbie Williams for pushing me to clarify this.

<sup>26</sup> Some B can fall outside of my network without being truly divergent from me, by there being some  $B' > B$  that is truly divergent from some  $B_{WG}' > B_{WG}$ . This situation is rather more complex and hence trickier to discuss, but fortunately we need not discuss the issue separately, because these options collapse *as far as smugness is concerned*. For if we are merely divergent, then any point of convergence between us *is* a part of my network and is truly divergent B' (proof follows), and I am not smug over *those*. (Note that this way of falling outside

First, consider that the bar for mere divergence is set incredibly low. All it requires is that there is *some* belief set that we both consider to be improvements on our own, or that some improved version of each of our belief sets consider to be an improvement, or that some improved version of these improvements consider to be an improvement, or that some improved version of these improved improvements... Ask yourself who you would expect to satisfy this condition. It is worth re-emphasising two points here: (i) the improvements are *actual* improvements by the lights of the standards one endorses, so if (e.g.) a stubborn person values open-mindedness and responsiveness to evidence, her stubbornness *per se* is no obstruction; and (ii) the information potentially available to each will include the other's way of looking at things. This latter point matters because we'd expect admirable outlooks to be suitably responsive to the existence of other admirable outlooks with divergent opinions (see Blackburn 1984: 201-2). As Blackburn points out (1999: 215-6), it is not even a given that those with abhorrent views should be so divergent from us. One might, with patience and skill, be able to talk around a member of the Taliban.<sup>27</sup>

One should not, I think, be tempted to think that such radical, irreparable conflict is possible if someone has similar standards to your own (recall mere divergence only requires the *possibility* of convergence, not its *inevitability*), unless, perhaps, they have an utterly different starting point, at which point the smugness worry immediately starts to look superficial. Once again, acknowledging this possibility strikes me as a kind of scepticism regarding one's own standards; and, again, I don't think it's coherent given the expressivist's other anti-sceptical commitments on this front. It is one thing to allow that there is much moral disagreement, and that this disagreement might be quite fundamental. But it is another thing altogether to allow that this moral disagreement might be *in-principle* irresolvable, regardless of how much more informed, sensitive, coherent, etc. the participants become. Could one plausibly maintain that this is the case for, e.g., a Kantian and a consequentialist?<sup>28</sup>

The standards we've been considering are platitudinous enough that it's difficult to even imagine what someone with different standards might look like. Could someone endorse *misinformation*, *insensitivity*, *immaturity*, *close-mindedness*, and *incoherence*? More seriously, I suppose, we

---

my network requires that B be in the sceptical scenario that two truly divergent belief sets are accessible to B via self-endorsed improvement.) I am thus only smug over B in this circumstance to the extent that I am smug over B', which I *am* truly divergent from. If it is permissible to be smug concerning truly divergent belief sets, like B', it is permissible to be smug over "unimproved" versions of those belief sets, like B. To see this, imagine that B' is a silly or abhorrent belief set, like those considered above and below.

Here is the relevant proof. Suppose for *reductio* that B' and some point of convergence for us, B<sub>C</sub>, are merely divergent. There is thus some point of convergence, B<sub>C'</sub>, for B' and B<sub>C</sub>. By the transitivity of self-endorsed improvement, B<sub>C'</sub> > B<sub>WG</sub>. But then B<sub>C'</sub> is merely divergent from B<sub>WG</sub>; so there is some point of convergence for the two, B<sub>C''</sub>. By the transitivity of self-endorsed improvement, B<sub>C''</sub> > B'. But then B<sub>C''</sub> is a point of convergence for B<sub>WG</sub>' and B', and there is *ex hypothesi* no such point of convergence.

<sup>27</sup> Of course, organisations like the Taliban also indoctrinate the well-meaning; but this is not, I submit, via a process of *actually* improving by the lights of the standards they endorse. If it *is*, then we run into the kind of sceptical worries I consider below.

<sup>28</sup> This raises the point that the kinds of consideration at play here are really problems for *anyone* who wishes to make sense of moral truth.

might imagine that *adherence to the Bible* or something outweighs all other standards for some agent. Again, it's far from obvious that someone like this really would be truly divergent, at least if they endorse some sensible standards as well; but if they are, perhaps because they don't endorse any such standards, then the smugness worry becomes superficial once more.

We thus lay down a challenge to the would-be objector: to provide some compelling example of a case where someone is both plausibly truly divergent from me, and yet the Smugness Worry is not superficial. The mere possibility that the Weak Super-Stability Hypothesis might hold for me and not for others does not show this. I cannot show that this challenge cannot be met; but the burden lies squarely with my opponent.<sup>29</sup>

One final point. Imagine that someone did manage to meet this challenge; I do not think this at all likely, but imagine it is so. Is this game over for the expressivist? I think not. For what we would have, by my estimation, would be a compelling argument for moral relativism. And far from this being problematic, the Weak Super-Stability Hypothesis provides a way of understanding precisely this moral relativism; i.e., that there is more than one network, and which moral obligations an agent is subject to depends on the network of which she is a part. (Instead of weak super-stability within *my* network, the relativistic theory of moral truth would have to be weak super-stability within *a* network.) For various independent reasons I do not find moral relativism especially palatable, but that doesn't matter here; what matters is that, if the case can be made, then we are actually *well-situated* to accommodate the relevant lesson.<sup>30</sup>

#### 5.4 Moral Truth as Weak Super-Stability

I have argued that the moral truths are epistemically constrained, and in particular for the Weak Super-Stability Hypothesis: that the moral truths are all and only those propositions that are weakly super-stable at  $B_{WG}$  (i.e., that there is some  $B \geq B_{WG}$  such that  $p \in B$  and there is no  $B' > B$  such that  $p \notin B'$ ). This constitutes an explanandum.

How might we explain the Weak Super-Stability Hypothesis on either a deflationary or correspondence theory of truth? The short answer is that we can't, at least not without supplementation. The deflationist obviously has no resources to work with. Indeed, the deflationist is in particular difficulty here, given that we saw above that the expressivist has to make room for

---

<sup>29</sup> It should be apparent how all this lends support to the contention of §5.2.4, that there is no true divergence among those  $B > B_{WG}$ , since not only is the bar for mere divergence already very low, but any such  $B$ s are further constrained by being accessible from the same (sensible) starting point: mine!

<sup>30</sup> This kind of relativism would open up a large number of choice-points for the expressivist. For instance, for those for whom different networks are accessible via self-endorsed improvement (e.g., those merely divergent from me who nonetheless fall outside my network – see fn.26) it may be *indeterminate* which network they are a part of, and thus the truth of those propositions on which those networks disagree may be indeterminate *for them*. On this line, such a possibility looks less like a sceptical scenario; so *perhaps* the expressivist can allow that she herself is in such a scenario. This is all very interesting, but I think we should leave proper investigation until the relativist's case has been made.



factually defective discourse, and the deflationist faces notorious problems on this front.<sup>31</sup> (This is a new “indirect” argument to the formerly-common view that there is a tension between expressivism and deflationism.)<sup>32</sup> The correspondence theory has the correspondence relation, but it is difficult to see what help this might be. We need some kind of *conceptual* entailment such that any weakly super-stable moral belief will stand in this substantive relation to some suitable moral truthmaker, in some way that is compatible with expressivism’s anti-realist motivations (§4.2.1). It’s not at all obvious how this might go.

Of course, we might offer some *other* explanation of the Weak Super-Stability Hypothesis that is consistent with one or other of these theories. For instance, we might try to formulate some epistemically constrained theory of the moral *facts*. I have no idea what such a proposal might look like, however.

My proposal, which I think is the natural proposal, is that moral truth is epistemically constrained; that it consists in some anti-realist truth property that predicts and explains the Weak Super-Stability Hypothesis. A full defence of this proposal will argue that this is the *best* explanation; but at this stage it wins by default because it is unclear what the alternative explanations could be.

My theory, as the reader will no doubt have anticipated, is roughly that moral truth consists in weak super-stability. But this proposal immediately requires finessing. After all, I have argued that all and only those moral propositions weakly super-stable at  $B_{WG}$  are true, and this is consistent with some proposition being weakly super-stable at some other belief set and yet untrue (or not weakly super-stable and yet true), but we do not want my (present) belief set appearing in our theory as to the nature of moral truth. That would be smugness pushed to the point of hubris; besides, it would presumably engender some kind of relativism, and it looks like we would each have a different theory of the nature of moral truth.

Fortunately, there is a simple fix. Recall that all and only the propositions that are weakly super-stable at  $B_{WG}$  are weakly super-stable *within my network*. The expressivist thus ought to hold that moral truth consists in weak super-stability *within her network*, where this is to be read *de re* and not *de dicto*. What matters is weak super-stability within the network of which I happen to occupy a node; the broader picture, given the argument of §5.3.3, being that we all occupy nodes within this network. Given this, neither me nor my present belief set appears in our theory of *nature* of moral truth (or, at least, my belief set doesn’t appear *qua* mine) – though that I occupy a node may be the only way I have of pointing at the relevant network.

Now, one worry that one might have about this proposal is that it is not built into the very nature of moral truth that there can be no moral dialethia (though there will, in fact, be no moral dialethia, given that I have argued that there will be no true divergence among those  $B > B_{WG}$ ). But

---

<sup>31</sup> There is a huge literature on this: see, e.g., Dummett (1959), David (1994), Field (1994b), Simmons (1999), Holton (2000), Greenough (2010).

<sup>32</sup> The reasoning here was very similar to that which drove the thought that expressivism is incompatible with minimalism about *truth-aptness*, which I suggested is mistaken (§4.2.3, especially fn.16 and the citations there).

to press this as an objection requires that this is a necessary constraint on our theory of moral truth, and it is rather difficult to see how this might be motivated within the expressivist's framework. After all, we cannot make antecedent appeal to the realm of moral truths that our moral judgements are seeking to represent. Instead, *consistency* seems to be as much one of the standards that we endorse as information, sensitivity, coherence, and so forth. And, as I have been stressing throughout, the expressivist cannot make antecedent appeal to the idea that any of these standards is an *actual* (as opposed to merely perceived) standard for improvement. It may, of course, be a particularly *central* standard; but that just means that we value it very highly, as we do many other central moral judgements. Nonetheless, we ought to be able to *make sense* of the possibility that such judgements may be in the wrong. But now notice that the situation in which both  $p$  and  $\neg p$  become strongly super-stable is just the one in which I abandon the standard of consistency through self-endorsed improvement; and this is how it should be, since this is just how I make sense of the possibility that I am wrong about consistency being a standard to which one's moral beliefs should be held. On the contrary, then, as I see it we have no grounds for insisting on this as a part of the very nature of moral truth; the expressivist therefore ought to reject it as a constraint.<sup>33</sup>

The resultant theory of truth is highly reminiscent of Wright's *superassertibility* (or *superwarrant*) and Lynch's *concordance* (§1.2.3), and that is to its credit. In particular, note that we are free to make reference, as Lynch does, to other propositions corresponding; for instance, in cashing out the standard of *information*. But the weak super-stability theory is a novel anti-realist theory of truth. I have argued that the expressivist is committed to the Weak Super-Stability Hypothesis on pain of incoherence; and that this theory offers a neat explanation of the Hypothesis.

## Conclusion

Together with the argument from Chapter 3 and 4, this completes my direct, metasemantically-driven argument for truth pluralism. Correspondence explains why certain actions performed by rational agents are successful; weak super-stability tells us which moral propositions we ought to believe. The nature of truth therefore varies for truthbearers in different domains; this is domain-restricted disjoint pluralism. With the view motivated, in Part III I turn to the objections that are typically raised against this kind of alethic pluralism. My considerations there will not rely on the particular domain-restricted truth properties I have advocated for in Part II, but are for the most part quite general to any disjoint inflationary pluralism.

---

<sup>33</sup> To sate those who find themselves strongly opposed here, for whatever reason, note that there are very nearby theories that do meet this constraint. For instance:  $p$  is true just in case  $p$  is weakly super-stable within my network and it's not the case that  $\neg p$  is weakly super-stable within my network. Indeed, from certain points of view, this proposal even has some advantages over the weak super-stability theory, though I think that, on balance, the straightforward weak super-stability proposal is to be preferred. I resist setting this out for reasons of space.

---

---

PART III

LOGIC, LIARS, AND LOGICAL FORM

---

---

## CHAPTER 6

# LOGIC, LOGICAL FORM, AND THE DISUNITY OF TRUTH

---

‘I think this is a good – in other words, a *prima facie* awkward! – question, but I think it’s a version of a very old question: one that is not a question that is triggered by pluralism but was already there for monism... It’s an old and teasing issue. But whatever the solution, it’s going to be something that lets us say the right thing. And the right thing to say is that what makes the conjunction true is that the first conjunct is true AND the second conjunct is true! That’s the right answer. We have to be allowed to use conjunction (or negation, or any other truth-functional operator) in characterizing the truth-conferrer. If you are not allowed to do that, you are lost.’

Wright (2013: 134-5).

In Part II, I argued for a variety of domain-restricted inflationary pluralism: that truthbearers with different subject matters are apt for different inflationary truth properties. Disjoint inflationary pluralists are meant to face a number of embarrassing problems the moment truthbearers apt for different properties are “mixed” together. There’s a family of problems here: the Problem of Mixed Inferences, of Mixed Compounds, of Mixed Atomics, and of Mixed Generalisations. Of these, the former two have received considerably more attention than the latter. In Part III, I take these problems in turn, addressing the first two here, Mixed Atomics in the next chapter, and Mixed Generalisations in the last.

It is the job of a comprehensive metaphysics of truth to tell us what the nature of truth consists in for any truthbearer. A proper part of this project is saying what truth consists in for truth-functional logical complexes; that is, for complex sentences whose status with regards to truth is determined entirely by the truth of their components. For the purposes of this chapter, I’m going to assume that ‘not’, ‘and’, ‘or’, ‘if... then’, and ‘neither... nor’ all voice truth-functional operators in this sense, symbolised by ‘ $\neg$ ’, ‘ $\&$ ’, ‘ $\vee$ ’, ‘ $\rightarrow$ ’, and ‘ $\downarrow$ ’.<sup>1</sup> Monists, of course, are ideologically committed to giving precisely the same answer in this subproject as they are in any other. Pluralists are not so committed.

The Problems of Mixed Inferences and Compounds identify roles for truth that the pluralist’s restricted properties are unable to play precisely in virtue of their restricted application. They are thus *taken to be* arguments in favour of postulating an unrestricted truth property, which is

---

<sup>1</sup> That is, with the truth-tables FT, TFFF, TTTF, TFFT, and FFFT, respectively. Given §5.1.4, my official view is that (first-order) moral negations are apt for weak super-stability.

the constitutive commitment of monism.<sup>2</sup> This need not be incompatible with pluralism, since one can be a moderate pluralist: these postulate an unrestricted truth property,  $T_U$ , which stands in a one/many relationship to the restricted truth properties,  $T_1, T_2$ , etc., such that the latter are *versions* of the former, in some sense.

But in fact the literature on the Problems of “Mixed” Inferences and Compounds is driven by a number of subtle confusions, or so I will argue. The Problems, I show, are not really “mixing” problems at all, but are just particular instances of perfectly general constraints on any metaphysics of truth; *structural* constraints – constraints on the *relations between* the truth of different truthbearers – that the monist is not guaranteed to satisfy *just in virtue of being a monist*. On the contrary, in Section 1, I will show that the monist is, if anything, at a methodological *disadvantage* here, since her ideological commitment places a further constraint on her metaphysics of truth to which the pluralist’s is not subject.

In Section 2, I propose a metaphysics for truth for complexes that is neutral with regards to the metaphysics of truth at the atomic level. My *form-restricted pluralism* suggests that complex truthbearers are themselves apt for distinct truth properties that are grounded in the truth properties of their components. This metaphysics satisfies the constraints of Section 1, and as such resituates the dispute between monism and pluralism to the atomic level, where it belongs. Nonetheless, this is a major dialectical victory for pluralism, since it straightforwardly dissolves two of the most influential objections to the theory. I also respond to objections.

## 6.1 “Mixing” Problems?

### 6.1.1 The Problem of “Mixed” Inferences

Suppose that truth property  $T_1$  is relevant for truthbearers in set  $D_1$  and that  $T_2$  is relevant for truthbearers in  $D_2$ , and suppose further that (1) is in  $D_1$  and (2) is in  $D_2$ . Consider, then, the inference that we’ll call *Cat*:

***Cat***

(1) Ajax is a household pet.

(1→2) If Ajax is a household pet, then eating Ajax is wrong.

Therefore:

(2) Eating Ajax is wrong.

---

<sup>2</sup>The concept/property distinction has not always been clearly in mind in discussions of these problems. Some presentations (e.g., Tappolet (1997, 2000), Küne (2003), Lynch (2009: 54-9)) are either explicitly or more charitably read as arguments in favour of *concept* monism, though others (Lynch (2009: 62-7), Cook (2011)) are definitely not. For my purposes here I’m happy to concede concept monism, and focus on the Problems as they matter for our metaphysics.

As an instance of *modus ponens*, **Cat** is valid. But validity is necessary *truth*-preservation. An argument is valid just in case: necessarily, if the premises are true, then the conclusion is true. But, according to the pluralist, the truth of (1) consists in  $T_1$  and the truth of (2) consists in  $T_2$ . Supposing, then, that (1) can be  $T_1$  without being  $T_2$  and (2) can be  $T_2$  without being  $T_1$ , it looks like **Cat** should be sound, by the pluralist's lights, despite there being *no* (relevant) property in common between the premise (1) and the conclusion (2). How, then, is the pluralist to characterise validity, if there is no single property that is necessarily preserved in valid inference?<sup>3</sup>

Some notes on this. (i) What if the truth properties are not independent in the way the argument suggests? For instance, perhaps any truthbearer that corresponds is weakly super-stable. We can get some version of the argument going as long as it's not the case that having any one of the pluralist's truth properties entails having all the others (and perhaps even then), which would risk collapsing pluralism back into monism, so we can set aside this possibility here. (ii) The pluralist *might* be able to formulate an extensionally adequate characterisation of *validity* using one or other of her truth properties. For instance, if (1) and  $(1 \rightarrow 2)$  are both weakly super-stable, then presumably (2) will be weakly super-stable too. But this would be to get the right result for the wrong reason; that is why I was careful to mention *soundness* in the presentation of the problem. That is, we want our characterisation of validity to be such that the inference is sound when (1) and (2) have *different* properties, and given the relevant independence of the truth properties, then as soundness is a species of validity, this means we cannot characterise validity in terms of a *single* property. (iii) The pluralist might simply opt out of the semantic account of validity, in favour of, say, a proof-theoretic account. Cotnoir says that 'pluralism ought to be consistent with a semantic account of validity as well' (2013b: 566), but I don't see where the support for this claim is. We might think of the fates of these two views as intertwined: e.g., that an argument for pluralism is an argument for the proof-theoretic account of validity too. Nonetheless, it is no doubt *better* for the pluralist if she can be neutral on this front.

Tappolet (1997, 2000) suggests that the Problem presents the pluralist with a trilemma: either (a) deny that mixed inferences can be valid; (b) allow that there *is* some truth property that is shared by the premises and conclusion in a sound argument (so, which *can* be shared by the premises and conclusion in a *valid* argument); or (c) reject the semantic account of validity. And the literature has pretty much followed her lead here. For two decades now, this has been guiding the dialectic.<sup>4</sup>

It is, I am convinced, a false trilemma. Assuming that rejecting (a) and (c) lands one on horn (b) is to assume that the semantic account of validity presupposes that there is a single *property* that

<sup>3</sup> See, e.g., Williamson (1994a), Sainsbury (1996), Tappolet (1997, 2000), Beall (2000), Lynch (2001, 2004, 2009), Pedersen (2006), Wright (2013), Cotnoir (2013b), and Stollo (2017).

<sup>4</sup> Tappolet thinks (b) renders pluralism redundant; the consensus now is that this is premature, given the possibility of moderate pluralism. (See also Pedersen 2010 and Wright 2012 for discussions of "the instability challenge".) Beall (2000) suggests that the pluralist endorse a many-valued logic; Pedersen (2006) a plural (re-)interpretation of logic (see fn.14 below). Cotnoir (2013b) proposes a novel semantic account of validity using an algebraic semantics. While I'm sympathetic to Cotnoir's proposal on its own terms, I think it – and the other suggestions – are simply unnecessary. Truth pluralism, I'll argue, leaves logic untouched.

can be shared by the premises and conclusion of a valid argument. But I do not think it presupposes this at all.

Presentations of this problem are often fixated on the idea that “necessary truth-preservation” requires that there is a single thing – truth – that is preserved from premises to conclusion:

‘If there is more than one property of truth... then there is no single property being preserved from premises to conclusion in [a mixed] argument.’ (Lynch 2004: 388)<sup>5</sup>

‘...for the strong alethic pluralist, no single property of truth would be preserved from premises to conclusion.’ (Wright 2005: 9)<sup>6</sup>

‘...what is the property that is preserved from premises to conclusions if they don’t share a unique truth-property?’ (Caputo 2012: 854)

‘The problem of mixed inferences... can be solved as there is a single truth property that is preserved across valid inference.’ (Edwards 2013c: 116, fn.8)

‘...if one accepts that there is more than one kind of truth, then, when an inference proceeds from premises that differ in the kinds of truth they possess, one must explain what sort of truth, if any, the conclusion has inherited from the premises.’ (Barnard & Horgan 2013: 192-3)

‘...there cannot be a single truth property that is transmitted from premises to conclusion, and the apparent validity of the inference cannot be accounted for in terms of necessary truth preservation.’ (Stollo 2017)

Indeed, often little *more* is said in order to generate the appearance of a problem. First, then, a cautionary point that shouldn’t need to be made, but probably does. Consider the concept, <preservation>. This is a *diachronic* concept: things are preserved across time. <Validity>, however, is a *synchronic* concept. Arguments are valid at a time. What’s the upshot? When we say that validity *is* “necessary truth-*preservation*”, we are talking *metaphorically*. Validity does not happen across time, so validity does not literally consist in something being preserved at all. Similar points can be made regarding “inheritance” and “transmission”. Consider also that there can be 0-place valid arguments with tautologous conclusions or valid arguments where the premises *cannot* be true together, so *a*

---

<sup>5</sup> See also Lynch (2006: 64; 2008: 125; 2009: 63).

<sup>6</sup> See also Wright (2010: 271).

*fortiori* nothing is “preserved” from premises to conclusion. These cases bring out the oddity of this way of presenting the argument, for *everyone* thinks that, in these cases, there is no single property that is “preserved”.<sup>7</sup> We should not be fooled by the metaphor alone into thinking that validity literally requires the preservation of some (single) thing.

No, “necessary truth-preservation” is just a nice way of talking about the principle that: necessarily, if the premises of the argument are true, then the conclusion is true.<sup>8</sup> What constraint, exactly, does this principle put on our *metaphysics* of truth? Does it really require that, when sound, the premises and conclusion *share* a property? Well, let me extract a constraint that it definitely *does* enforce, for our mixed inference **Cat**:

*Semantic Validity Constraint (SVC)*

If the truth of (1) consists in F, the truth of  $(1 \rightarrow 2)$  in G, and the truth of (2) in H, then F, G, and H must be such that: necessarily, if (1) is F and  $(1 \rightarrow 2)$  is G, then (2) is H.

Or generally:

*General Semantic Validity Constraint (SVC\*)*

For a valid argument from premises  $\{A_1, \dots, A_n\}$  to B, if the truth of  $A_1$  consists in  $F_1, \dots$ , the truth of  $A_n$  in  $F_n$ , and the truth of B in G, then  $F_1, \dots, F_n$ , and G must be such that: necessarily, if  $A_1$  is  $F_1, \dots$ , and  $A_n$  is  $F_n$ , then B is G.<sup>9</sup>

For simplicity, let’s focus on the particular constraint, SVC.

The first thing to note about SVC is that it is a *perfectly general constraint on the metaphysics of truth*. Monists and pluralists alike must satisfy it.

The second thing is that there is nothing about postulating an identity between F, G, and H, as the monist does, that makes satisfying SVC any easier. This is obvious. There is absolutely no reason to think that, in general, both (1) and  $(1 \rightarrow 2)$  exemplifying some property  $F^*$  guarantees that (2) will exemplify that property. Consider, for example, the property of *being a premise in Cat*. (1) and  $(1 \rightarrow 2)$  exemplify this property; (2) does not.

So far as this constraint is concerned, then, the monist *qua monist* is at no immediate advantage. Indeed, the monist is plausibly at a methodological *disadvantage* here, since her ideology places a *further* constraint on her metaphysics of truth: the monist *has* to postulate an identity between

---

<sup>7</sup> Stollo (2017) shows that he is aware of these points.

<sup>8</sup> This principle features the word ‘true’ twice, so one worry concerns *equivocation*: how can the pluralist account for the *semantic* unity of these uses given the underlying *metaphysical* diversity? The pluralist’s answer ought to be that it voices a single *concept*. One might then worry that conceptual monism entails property monism; but that is a different argument (see §1.4.2 and Chapter 8).

<sup>9</sup> Sher (2013: 158-9) thus gets much closer to the right construal of the Problem. With it so understood, it is obvious that it cannot be solved simply by endorsing conceptual monism, as is suggested by Wright (2013: 132-3).



F, G, and H. Moreover, in discussion of the Problem of Mixed Compounds in the next subsection, we'll see that various monists, correspondence theorists chief among them, will plausibly *struggle* to satisfy SVC. Therefore, far from *requiring* monism, for all we've seen so far this constraint may be *more* difficult to satisfy if one is a monist.

If SVC\* is the only constraint enforced by a semantic account of validity, then worries that trade on there being 'no single property preserved by the inference' (Lynch 2009: 63) are ill-founded. Indeed, it is tempting to think that most of the intuitive pull of the Problem of "Mixed" Inferences comes from taking the "preservation" metaphor a little too literally. But perhaps SVC\* is not the strongest such constraint. I think that it is, but rather than considering and rejecting other proposals at this stage, it will be more productive to propose a metaphysics that satisfies this constraint without postulating an unrestricted truth property. With this hand, we can ask what is *missing*, as far the semantic account of validity is concerned. It is incumbent upon the objector to argue that something more is needed.

One *might* suspect, on slightly different grounds, that the pluralist will struggle to satisfy SVC because she postulates diversity between F and H, the thought being that identifying some property G such that SVC is satisfied might therefore be more difficult. This is essentially the charge at the heart of the Problem of Mixed Compounds, to which we now turn.

### 6.1.2 The Problem of "Mixed" Compounds

Consider the following truth-functional compounds:

- ( $\neg$ 1) Ajax is not a household pet.
- (1&2) Ajax is a household pet and eating Ajax is wrong.
- (1 $\vee$ 2) Ajax is a household pet or eating Ajax is wrong.
- (1 $\rightarrow$ 2) If Ajax is a household pet, then eating Ajax is wrong.
- (1 $\downarrow$ 2) Neither is Ajax a household pet, nor is eating Ajax wrong.

As mentioned at the beginning of the chapter, any comprehensive metaphysics of truth owes us an account of what the truth of such compounds consists in. The following constitute schematic extensionality constraints on our theory (using single-quotes here as a device for taking a declarative sentence to the relevant primary truthbearer):

#### *Truth-Functional Extensionality Constraints*

- ' $\neg$ p' is true  $\leftrightarrow$   $\neg$ ('p' is true).
- 'p & q' is true  $\leftrightarrow$  ('p' is true & 'q' is true).
- 'p  $\vee$  q' is true  $\leftrightarrow$  ('p' is true  $\vee$  'q' is true).

‘ $p \rightarrow q$ ’ is true  $\leftrightarrow$  (‘ $p$ ’ is true  $\rightarrow$  ‘ $q$ ’ is true).

‘ $p \downarrow q$ ’ is true  $\leftrightarrow$  (‘ $p$ ’ is true  $\downarrow$  ‘ $q$ ’ is true).

To see the *prima facie* worry for the pluralist, consider the conjunction (1&2). The truth of (1) consists in  $T_1$ , the truth of (2) consists in  $T_2$ . What, then, does the truth of (1&2) consist in? There’s no guarantee that the conjunction *as a whole* will be either  $T_1$  or  $T_2$  when (1) is  $T_1$  and (2) is  $T_2$ , especially given that (1) might be  $T_1$  and not  $T_2$ , while (2) might be  $T_2$  and not  $T_1$ .<sup>10</sup>

The problem with this rather schematic way of putting things, of course, is that whether or not the conjunction as a whole can fail to be either  $T_1$  or  $T_2$  when (1) is  $T_1$  and (2) is  $T_2$  depends entirely on the nature of the truth properties involved; it doesn’t simply follow from the mere fact that  $T_1$  and  $T_2$  are distinct. (For instance, if (1) corresponds then it is plausibly weakly super-stable, and if (2) is also weakly super-stable, then the conjunction may be too.) But still, let’s grant this for the sake of argument, so the truth of the conjunction cannot consist in  $T_1$  or  $T_2$ .

What do the extensionality constraints tell us about the metaphysics of truth? Here, some theorists have once again suggested that identity flows immediately, at least when the compound in question is a conjunction:

‘Mixed conjunctions need to be true in a further way... But then each conjunct needs to be true in the same way. This is what follows from the truism that a conjunction is true if and only if its conjuncts are true.’ (Tappolet 2000: 385)

‘...the principle that a conjunction is true if and only if its conjuncts are true is not negotiable... So application of the conjunction platitude to such cases [like (1&2)] would be open to the charge of equivocation if the conjuncts were not true in the *same* sense.’ (Künne 2003: 453)

‘A conjunction is true if and only if the conjuncts are true, and further, the conjunction should be true in the same way as the conjuncts are true. Hence neither  $T_1$  nor  $T_2$  is the appropriate notion of truth for this conjunction: if the conjunction were  $T_1$ -true, then both [(1)] and [(2)] would need to be  $T_1$ -true, but there is no general guarantee that [(2)] is  $T_1$ -true (and, similarly, if the conjunction were  $T_2$ -true, then both [(1)] and [(2)] would need to be  $T_2$ -true, but there is no general guarantee that [(1)] is  $T_2$ -true). Thus there must be some third truth property  $T_3$  in virtue of which statements from the combined discourse are true.’ (Cook 2011: 626)<sup>11</sup>

---

<sup>10</sup> See, e.g., Williamson (1994a), Tappolet (2000), Lynch (2004, 2009), Edwards (2008, 2009), Cotnoir (2009), Cook (2011), Wright (2013).

<sup>11</sup> It is interesting that Tappolet, Künne, and Cook all suggest slightly different constraints. Tappolet says that, if the truth of ‘ $p \& q$ ’ consists in  $F$ , then the truth of ‘ $p$ ’ consists in  $F$  and the truth of ‘ $q$ ’ consists in  $F$ . Cook

But, in fact, the only constraints that we can extract from the extensionality constraints alone are the following. Where ‘!’ is a schematic marker for a truth-functional operator, if the truth of ‘p’ consists in F, the truth of ‘q’ consists in G, and the truth of ‘p ! q’ consists in H<sub>!</sub>, then:

‘¬p’ is H<sub>¬</sub> ↔ ¬(‘p’ is F).

‘p & q’ is H<sub>&</sub> ↔ (‘p’ is F & ‘q’ is G).

‘p ∨ q’ is H<sub>∨</sub> ↔ (‘p’ is F ∨ ‘q’ is G).

‘p → q’ is H<sub>→</sub> ↔ (‘p’ is F → ‘q’ is G).

‘p ↓ q’ is H<sub>↓</sub> ↔ (‘p’ is F ↓ ‘q’ is G).

To this, we can plausibly add a direction of explanation: ‘p & q’ is true, when true, not just in all and only the cases where ‘p’ and ‘q’ are true, but *because* ‘p’ and ‘q’ are true (Edwards 2008: 146–7). The ‘because’ here is the ‘because’ of constitutive explanation or grounding: a true conjunction *just is* a conjunction with two true conjuncts. It is tempting – as will be prominent in the metaphysics offered in the next section – to say: the truth of the conjunction *consists in* the combined truth of its conjuncts.

Again, these are *perfectly general constraints* to which any comprehensive metaphysics of truth is subject. And again, merely postulating an identity between F, G, and H<sub>!</sub> does not render the task of satisfying them any easier. A negation does not have every property that its negand does not have; nor a conjunction every property shared by its conjuncts; nor does a disjunction inherit every property possessed by one of its disjuncts.

Indeed, the monist faces a *prima facie* difficulty that we can call *the concatenation problem*. Given the identity she postulates between H<sub>&</sub>, H<sub>∨</sub>, H<sub>→</sub>, and H<sub>↓</sub> with her preferred truth property T<sub>U</sub>, she must explain why the concatenation of two truthbearers that are T<sub>U</sub> *in a conjunction or a disjunction* results in a complex that is T<sub>U</sub>, while their concatenation in a neither/nor construction does not; and likewise, why the concatenation of two truthbearers that are not T<sub>U</sub> results in a complex *that is* T<sub>U</sub> when the complex is a (material) conditional or a neither/nor construction, but not when the complex is a conjunction or disjunction. My point is not that this is some insuperable difficulty. It is that, whatever the monist’s explanation is, it cannot consist in *mere concatenation*. Not in any case (even for a conjunction), since even by the monist’s own lights, mere concatenation of two truthbearers that are T<sub>U</sub> is neither necessary nor sufficient for the resultant complex to be T<sub>U</sub>. The explanation, whatever it is, will require more than this.

---

(at least initially, but I think throughout) suggests the other direction: if the truth of ‘p’ (or ‘q’) consists in F, then the truth of ‘p&q’ consists in F. Künne’s claim, strictly, only amounts to something like: if ‘p&q’ is true, then the truth of ‘p’ consists in F iff the truth of ‘q’ consists in F. (To be fair to Künne, his concern here is with the concept <truth> and hence the charge of equivocation.)

One might be tempted to “go recursive”, citing, e.g., the authority of Tarski: a conjunction is true if and only if its conjuncts are true!<sup>12</sup> Right. But this is a *datum* that our theory has to predict and explain. The only way I can read this as telling us anything about the *nature* of truth is if we read it in accordance with the form-restricted pluralism I’ll articulate below. This results, as I’ll explain, in a radical disunity in the nature of truth; if everyone who endorses recursion here is happily committed to this radical pluralism then they’re being awfully quiet about it! (Or, more likely, they’ve simply not realised it.) But if this is *not* what is intended, then the recursive definition here amounts to nothing more than a repetition of the very constraint at hand, and repetition is not explanation.

That the task is non-trivial is foregrounded by the fact that the traditional monists may well *struggle* to satisfy these constraints:

‘it is not the case that p’ coheres  $\leftrightarrow$  ‘p’ does not cohere.

‘p or q’ is superwarranted  $\leftrightarrow$  (‘p’ is superwarranted or ‘q’ is superwarranted).

‘neither p nor q’ corresponds  $\leftrightarrow$  (neither ‘p’ corresponds nor ‘q’ corresponds).

None of these is obviously true. That a belief does not cohere with a set of beliefs does not guarantee that its negation will cohere with that set of beliefs (perhaps the beliefs do not lend support either way). Similarly, a disjunction may plausibly be superwarranted despite the fact that neither of its disjuncts are. We might have a warrant that one or the other of the disjuncts is true without having a warrant that one of them in particular is so. And the final one is just one of a number of embarrassing problems faced by the correspondence theory. Do we want to say that there are conjunctive facts, or disjunctive facts, or (shudder) *negative* facts that conjunctions, disjunctions, or negations correspond to?<sup>13</sup>

Of course, there are a multiplicity of responses the monist might give to these worries; but the fact that they need to give responses illustrates precisely the point I’m making. *Merely being a monist*, even of one of the widespread, “popular” varieties, does not guarantee that one’s theory of truth satisfies the relevant constraints.

Once we acknowledge this, it becomes apparent that the monist is once more at a tactical *disadvantage* here. For she is ideologically *constrained* to postulate an identity between F, G, and H, where the pluralist is not. Contrary to the (unargued) claims of Tappolet, Künne, and Cook, then,

---

<sup>12</sup> Thus one sometimes hears or reads talk of ‘the recursive strategy familiar since the age of Logical Atomism’ (Caputo 2012: 856; see, e.g., Lynch 2009: 86-91).

<sup>13</sup> The problem for correspondence theorists has been pointed out in this context by Edwards (2008) and Wright (2013: 134-5). Truthmaker maximalists have gone to extremes to avoid postulating negative facts, including postulating totality facts (Armstrong 2004) and absences (Martin 1996), or even denying that there are any negative truths (Mumford 2007). Given these kinds of worries, I simply do not understand Cotnoir’s (2009: 477-8) proposal that we “let” negations have the relevant truth properties of their negands and disjunctions the truth property or properties of their disjuncts (perhaps both!). As the correspondence case renders especially salient, whether or not these are plausible theories for the *negation* or *disjunction* as a whole is an open question that may well have a negative answer.

the raw extensionality constraints themselves do not legislate in favour of identity; indeed, being ideologically committed to monism merely makes the task of satisfying them that much more difficult.

When it comes to validity, the monist can *say* that it is her preferred truth property that is “preserved” across valid inference. But I have argued that this is not really what is at stake. The monist needs to show that she can satisfy SVC(\*); and *that* involves giving a satisfactory metaphysics of truth for complex truthbearers.<sup>14</sup> Both of the Problems of “Mixed” Inferences and Compounds have thus transpired to be instances of perfectly general constraints on the metaphysics of complex truth, which the monist too has to satisfy, and with regards to which her mere monism constitutes no progress.

Nonetheless, the pluralist does face some challenges that the monist does not. The first challenge here comes from *inappropriate truth properties*. Suppose that (1) is not  $T_1$ , but is  $T_2$ , and (2) is not  $T_2$ , but is  $T_1$ ; thus neither is true, though each possesses *a* truth property. Our metaphysics of truth for, e.g.,  $(1\&2)$ , needs to be such that  $(1\&2)$  is  $H_{\&}$  when (1) is  $T_1$  and (2) is  $T_2$ , but *not* when these properties are “reversed”. Likewise,  $H_{\downarrow}$  needs to be such that  $(1\downarrow 2)$  is  $H_{\downarrow}$  when (1) is not  $T_1$  and (2) is not  $T_2$ , even if (1) is  $T_2$  and (2) is  $T_1$ .<sup>15</sup> The second challenge comes from the sheer fact that the pluralist tolerates diversity in the nature of truth for the components of a complex: F and G in the above schemas. Suppose that some truthbearer (3), like (1), is in  $D_1$ , and (4), like (2), is in  $D_2$ . One might wonder what property  $H_{\&}$  can satisfy all three of:

$$\begin{aligned} (1\&2) \text{ is } H_{\&} &\leftrightarrow ((1) \text{ is } T_1 \ \& \ (2) \text{ is } T_2). \\ (1\&3) \text{ is } H_{\&} &\leftrightarrow ((1) \text{ is } T_1 \ \& \ (3) \text{ is } T_1). \\ (2\&4) \text{ is } H_{\&} &\leftrightarrow ((2) \text{ is } T_2 \ \& \ (4) \text{ is } T_2). \end{aligned}$$

Note, however, that while the pluralist is certainly committed to giving some metaphysics of truth for each of the different conjunctions, she is *not* antecedently committed to thinking that all conjunctions are true in the same way. That’s the monist’s ideology. The pluralist is simply not committed to the identity of the above “ $H_{\&}$ ” properties.

Of course, if she does postulate diversity here, it is incumbent on her to show that this is tolerable. But that is just part of giving a satisfactory metaphysics of (complex) truth. These challenges are, then, just that: challenges. Whether or not they can be met is at this point an open question. I’ll confront the issue directly by presenting a satisfactory and pluralist-friendly

---

<sup>14</sup> This point also scuppers other pluralist attempts to solve the Problem of “Mixed” Inferences. For instance, Pedersen’s (2006) proposal requires reconstruing validity as: necessarily, if the premises have properties *from among the truth properties*, then the conclusion must have a property *from among the truth properties*. But which of the truth properties does  $(1\rightarrow 2)$  have?

<sup>15</sup> Again, it is no mere truism that truthbearers in one domain *can* possess truth properties relevant for other truthbearers; nor that they can do so (if they can) without possessing their own truth properties. It very much depends on the metaphysics of the truth properties involved. But I continue to set this aside.

metaphysics of complex truth in the next section. In this section, however, I have argued that the Problems of “Mixed” Inferences and Compounds are just instances of perfectly general constraints on the metaphysics of truth, and that the monist *qua monist* is no better situated to satisfy them than the pluralist is.

### 6.1.3 Against Moderate Pluralism

It’s worth pausing to note that being a moderate pluralist constitutes no advantage on this front. Moderate pluralists maintain that there is an unrestricted truth property,  $T_U$ , which is a genus, determinable, functional role property, or disjunctive property, of which the domain-restricted truth properties are the species, determinates, realisers, or disjuncts.<sup>16</sup> I call this the “One/Many Gambit”. The moderate pluralist will want to say:

$$\begin{aligned} \neg p \text{ is } T_U &\leftrightarrow \neg(p \text{ is } T_U). \\ p \ \& \ q \text{ is } T_U &\leftrightarrow (p \text{ is } T_U \ \& \ q \text{ is } T_U). \\ p \ \vee \ q \text{ is } T_U &\leftrightarrow (p \text{ is } T_U \ \vee \ q \text{ is } T_U). \\ p \rightarrow q \text{ is } T_U &\leftrightarrow (p \text{ is } T_U \rightarrow q \text{ is } T_U). \\ p \downarrow q \text{ is } T_U &\leftrightarrow (p \text{ is } T_U \downarrow q \text{ is } T_U). \end{aligned}$$

Where she’ll explain the components’ being  $T_U$  in at least some cases in virtue of their possessing some restricted truth property,  $T_1, T_2$ , etc.

I’ve already argued that merely postulating an identity here doesn’t help. But the moderate pluralist faces a further difficulty. Supposing that  $T_U$  is a genus, determinable, functional role, or disjunctive property, we can ask *which* species, determinate, realiser, or disjunct the complex truthbearers on the left-hand side are supposed to possess.  $T_1$  and  $T_2$  are meant to be species, determinates, realisers, or disjuncts of  $T_U$ , but the whole point of the “Problem of Mixed Compounds” is that  $T_1$  and  $T_2$  look ill-suited to play this role. If not, there’d be no problem and  $T_U$

---

<sup>16</sup> The first three are the philosophically familiar one/many relations. Correspondence pluralists are probably best interpreted in terms of genus/species (Pedersen & Wright (2013a: 7) seem to agree). Edwards (2011, 2013c, fc) calls his theory “simple determination pluralism” and says the restricted properties “determine” the unrestricted property, but stops short of saying the relation is the determinable/determinate one. Lynch (2001, 2004, 2005, 2006; see also Pettit 1996) defended the “alethic functionalist” theory that  $T_U$  is the second-order property of having a property that realises the “truth-role”, where the “role” is identified via platitudes about truth and different restricted properties realise this role in different domains. “Alethic disjunctivism”, where  $T_U$  is defined in terms of disjunction of the domain-restricted properties, perhaps conjoined with their relevant domains, is explored and defended in, e.g., Pedersen (2006, 2010), Edwards (2012b), and Pedersen & Wright (2013b). I consider Lynch’s more recent “manifestation functionalism” below. Each view suffers difficulties on its own terms, but there may be other, more plausible, versions. I settle here for showing that they are unmotivated by the Problems under consideration. This is significant: given domain-restricted pluralism, the “mixing” problems are often cited by way of motivating moderate pluralism.

would be redundant, at least as far as this problem is concerned. Furthermore, some complexes are meant to be  $T_U$  despite none of their components being either  $T_1$  or  $T_2$ .

Perhaps the complexes possess some *other* species, determinate, realiser, or disjunct of  $T_U$ ; perhaps one somehow “composed” or “constructed” from the component’s species, determinates, realisers, or disjuncts. Maybe. But that requires that we identify some *other* truth property that the complex possesses when true (perhaps one “composed” or “constructed” from the component’s truth properties), which is just the task of giving a metaphysics of truth for logical complexes. Again, if we had that then there’d be no problem, and  $T_U$  would be redundant. So, that’s no progress on the problem.

Relatedly, perhaps the complexes possess  $T_U$  despite *not* possessing any particular species, determinate, realiser, or disjunct of it, but nonetheless in virtue of its components possessing their relevant truth properties (somehow). But this is a very odd suggestion indeed. Something cannot typically be of a genus without being of a species of that genus (a bear that is no particular species of bear?), nor possess a determinate without possessing a determinate of that determinate (a coloured object that has no colour; a red object that is no shade of red?), a functional role property without having a realiser of that role (a kidney that doesn’t play the role of a kidney?),<sup>17</sup> nor a disjunctive property without satisfying some disjunct (something that is *either red or a cow* despite being neither red nor a cow?). Many of these sound completely conceptually confused.

The possible exception is in the case of determinable/determinate: perhaps if we glue together something that is scarlet with something that is crimson, we’ll have something that is red (all over) despite being no *particular* shade of red (all over). But even if this is right, it is of no help. It provides a model for conjunctions, but this model goes out the window once we recall the concatenation problem: mere concatenation of two truthbearers that are  $T_U$  is neither necessary nor sufficient for the complex as a whole to be  $T_U$ . The complex’s possession of  $T_U$  would still need explaining.

Lynch realised that his erstwhile alethic functionalist view faced the problem of saying which property realised  $T_U$  for complex truthbearers, and thus abandoned that view.<sup>18</sup> His ingenious response was to make up a brand new one/many relation – manifestation – which allows that something can possess the one despite not possessing any of the many (Lynch 2009: 74). Indeed, the manifestation relation is *reflexive*, so *every* property self-manifests. According to Lynch, a property F manifests a property G just in case it is a priori that the conceptually essential features of G are a subset of the features of F. Lynch then thinks that  $T_U$  can be manifested in  $T_1$ ,  $T_2$ , and so on. He thinks of his view as a variety of functionalism still, because he thinks the conceptually essential features of  $T_U$  delineate a functional role; but that doesn’t matter for our purposes. In fact, the

---

<sup>17</sup> Perhaps a malfunctioning kidney is a kidney that doesn’t play the role of a kidney. But the functionalist presumably does not want to say that the complex’s position is analogous to that of a malfunctioning kidney. (The analogy here is difficult to even make sense of, but that is the functionalist’s problem.)

<sup>18</sup> E.g., Lynch (2009: 65-7).

proposal has all kinds of problems on its own terms.<sup>19</sup> But in any case, since it is a priori that any property F's conceptually essential features are a subset of its own features, every property self-manifests. (All the time. They even self-manifest while they're busy being manifested in other properties.) And this, Lynch thinks, allows him to divide the truths – those truthbearers with  $T_U$  – into the “unplain” truths, for whom  $T_U$  is manifest in some other property, and the “plain” truths, for whom  $T_U$  exclusively self-manifests. Mixed compounds can then simply be plain truths! Problem solved. (Lynch 2009: 86-91.)

If only things were so easy. Consider again the concatenation problem. Nothing in Lynch's metaphysics tells us why  $T_U$  chooses to self-manifest when a conjunction's components are  $T_U$ , but not when a neither/nor construction's are so; nor why it chooses to self-manifest when a neither/nor construction's components are not  $T_U$ , but not when a conjunction's are so. All it tells us is that  $T_U$  sometimes chooses to self-manifest, and sometimes chooses not to; and in all and only the right cases. But then Lynch has only succeeded in labelling the phenomenon we want explaining.

Of course, we can imagine Lynch *saying* “ $T_U$  self-manifests for the conjunction because conjunctions are true if and only if their conjuncts are true”. But *this is not an explanation*; it is an explanandum! We want the property that a conjunction possesses when true to be a property it possesses just in case its conjuncts are true; but just *saying* that the property one chooses to play this role satisfies this constraint (because it “self-manifests” just in case this is so) is just repeating the very constraint at hand. Again, repetition is not explanation.

Lynch tries to instigate a tighter connection between the self-manifesting  $T_U$  of a mixed compound and the unplain truth or otherwise of its components by endorsing the following “weak grounding principle”:

‘There can be no change in the truth-value of a compound proposition without change in the truth-value of *some* atomic propositions. The truth-value of compounds supervenes on the truth-value of atomic propositions... What makes a compound proposition plainly true? Given the weak grounding principle, compound propositions are plainly true if their truth-value is grounded. That is, if their truth-value supervenes on the truth-value of propositions which are either compound and grounded or atomic...’ (Lynch 2009: 90-1)

For one thing, supervenience is too weak for explanation.<sup>20</sup> But, in any case, *nothing* about Lynch's metaphysics predicts the weak grounding principle. It's simply a promissory note to which Lynch is voicing his commitment, and which therefore does nothing to render the convenient self-

---

<sup>19</sup> E.g., Tappolet (2010: 1195-7), Pedersen & Edwards (2011), Shapiro (2011), Caputo (2012), Edwards (2012a), Horton & Poston (2012), Pedersen (2012), Newhard (2013, 2014, 2017), Pedersen & Wright (2013b). See Lynch (2012, 2013b) for some replies.

<sup>20</sup> Caputo (2012) and Pedersen & Wright (2013b) also raise worries about the explanatory status of manifestation.



manifesting behaviour of  $T_U$  any less mysterious. We’re all committed to the weak grounding principle: it’s a *constraint* we need to show our metaphysics meets, not a principle to which we are antecedently entitled.

Moderate pluralists are thus no better off when it comes to the Problem of “Mixed” Compounds; i.e., when it comes to giving a metaphysics of complex truth. And in turn, like traditional monists, they’re no better off when it comes to the Problem of “Mixed” Inferences. They can *say* that it is  $T_U$  that is “preserved” in valid inference, but the real challenge is providing a metaphysics of truth that satisfies SVC(\*). Since *Cat* contains a mixed compound,  $(1 \rightarrow 2)$ , the moderate pluralist has no solution to the Problem of “Mixed” Inferences until she has a solution to the Problem of “Mixed” Compounds.

## 6.2 Form-Restricted Pluralism

### 6.2.1 The Theory

We need to say what the truth of a compound consists in, when it is true; and we need to do so in such a way that the constraints extracted in Section 1 are satisfied. My proposal will, I think, strike some as utterly absurd and others as platitudinous and boring. Let me say this from the outset. My proposal is pluralistic: different complex truthbearers are apt for different properties. But (perhaps surprisingly) for each property that I say is relevant for some set of truthbearers  $S$ , *every* inflationist is committed to all and only the true members of  $S$  having that property. To this extent, my proposal is utterly non-revisionary. But it is “revisionary” in suggesting that this is what their truth *consists in* (though I scare-quote the revisionism, because I think many theorists may be committed to this). What is new is, I think, that I take the ensuing disunity in the nature of truth seriously. That is what gives form-restricted pluralism, on first contact, its paradoxical *platitudinous-yet-radical* feel.<sup>21</sup>

Complex truthbearers come in different orders of complexity. We will say a truthbearer is of the  $n^{\text{th}}$ -order just in case its highest-order component is of the  $(n-1)^{\text{th}}$  order, where atomics are  $0^{\text{th}}$ -order. Let ‘ $T_A$ ’, then, label the property that the truth of all atomics consists in. For the monist, ‘ $T_A$ ’

---

<sup>21</sup> Symptomatic of this is that Edwards (2008) proposes something that sounds very similar to my proposal; but when Cotnoir (2009) calls him out on it, Edwards (2009) disavows form-restricted pluralism. Instead, he maintains that he was only talking about truth *conditions*. He then identifies the relevant truth property for *any* logical compound with whatever property is relevant for truthbearers in the “logical domain” – i.e., for truthbearers *about* logic – about which he tells us nothing. This is on the one hand surprising and counter-intuitive: after all,  $(1 \& 2)$  is about Ajax and the moral status of eating him, not about logic. But it also *doesn’t solve the problem* until we’re told what the nature of this property is. This makes it all the more remarkable that Strollo (2017) tries to leverage this proposal into a solution to the Problem of “Mixed” Inferences (by saying that valid inferences are conditionals – with the conjunction of the premises as antecedent and the conclusion as consequent – that have this truth property), again without offering a single detail about what the nature of this property is. Until we are given a plausible story here these proposals are no proposals at all; we might as well say that the truth of the complex consists in something-or-other.

refers to correspondence, coherence, superwarrant, or what have you. For the pluralist, it abbreviates a disjunction: being either (in  $D_1$  and  $T_1$ ) or... or (in  $D_n$  and  $T_n$ ). The disjunctive aspect of this proposal is ultimately dispensable, but it will be easier to work with it for the time being. Further, let ‘ $T_{iii}$ ’ label the property that the truth of the relevant first-order complex consists in. Here, then, are my rather mundane theories as to the nature of these properties:

$$\begin{aligned} \forall p (\Gamma_{\neg 1}(\neg p) &\leftrightarrow \neg T_A(p)). \\ \forall p \forall q (\Gamma_{\& 1}(p \& q) &\leftrightarrow (T_A(p) \& T_A(q))). \\ \forall p \forall q (\Gamma_{\vee 1}(p \vee q) &\leftrightarrow (T_A(p) \vee T_A(q))). \\ \forall p \forall q (\Gamma_{\rightarrow 1}(p \rightarrow q) &\leftrightarrow (T_A(p) \rightarrow T_A(q))). \\ \forall p \forall q (\Gamma_{\downarrow 1}(p \downarrow q) &\leftrightarrow (T_A(p) \downarrow T_A(q))).^{22} \end{aligned}$$

That is, the truth of a first-order negation consists in *being a negation that negates a truthbearer that is not  $T_A$* , or just *negating a truthbearer that is not  $T_A$* ; similarly, the truth of a first-order conjunction consists in *being a conjunction of a truthbearer that is  $T_A$  with another truthbearer that is  $T_A$* , or simply *conjoining two truthbearers that are  $T_A$* ; and so on.

First, this metaphysics straightforwardly satisfies the constraints laid down by the Problems of “Mixed” Inferences and Compounds. The relevant truth property for  $(1 \rightarrow 2)$  is such that, if (1) is  $T_A$  and  $(1 \rightarrow 2)$  is  $T_{\neg 1}$ , then (2) must be  $T_A$ . This is consistent with the nature of truth consisting in  $T_1$  for (1) and  $T_2$  for (2), for any such  $T_1$  and  $T_2$ . Of course, no single property is *literally* preserved across the inference; but we should never have been tempted to think that any property was. The extensionality (and grounding) constraints are satisfied trivially.

Despite essentially reading the metaphysics off of the extensionality constraints, this account does not merely repeat the constraints, though it does make good on the intuition driving the thought that there is nothing more to say about the nature of truth of a conjunction besides the truth of its conjuncts. (And the analogous intuitions for the other compounds.) There is no concatenation problem for the form-restricted pluralist, because the truth of the different complexes consists in different properties.

It’s obvious, but it’s worth saying explicitly: the property of *being a conjunction of a truthbearer that corresponds (or whatever) with another that corresponds* is a distinct property from the property of *corresponding*. If this is your metaphysics of truth for complexes – as I imagine it implicitly is for many tempted by recursive accounts here – you ought to be interested in the ramifications of the resultant pluralism.

---

<sup>22</sup> The single-quotes here should strictly be understood as so-called “quasi-quotes”, where this is a metalinguistic device that allows us to refer to the form of an expression without referring to the symbols.

One may doubt that these properties are really *properties*, in a special sparse or natural sense. This does not concern me here. My claim is that this is what their truth consists in. If one would prefer to translate my talk of “properties” into other terms, feel free to do so.

These theories only cover first-order complexes of the relevant kinds of complex. We will need further truth properties for higher-order complexes. The proposal thus results in an “infinite proliferation” of truth properties (Cotnoir 2009). Given the dependence of a truth-functional complex on the truth of its components, this will ultimately always be cashed out in terms of  $T_A$ . For instance, for a truthbearer of arbitrary complexity, ‘ $p \rightarrow ((q \vee r) \& \neg(s \vee t))$ ’:

$$\forall p \forall q \forall r \forall s \forall t (T_{\rightarrow 4A \& 3V1 \neg 2V1}(p \rightarrow ((q \vee r) \& \neg(s \vee t))) \leftrightarrow (T_A(p) \rightarrow ((T_A(q) \vee T_A(r)) \& \neg(T_A(s) \vee T_A(t)))).$$

The “complexity” of the truth property – i.e., the complexity of the definition on the right-hand side of the biconditional that tells us the nature of the property – thus mimics the complexity of the truthbearer itself.

(Note, then, that as I am cashing this out, different truth properties will be relevant for truthbearers of the same kind of the same order of complexity, depending on the logical form of their components. For a similar reason, I am happy to allow that, granting domain-restricted pluralism, different truthbearers of the same order of complexity are apt for different properties – characterised in terms of either  $T_1$ ,  $T_2$ , or both – depending on the *contents* of their components. For ease I stick with using ‘ $T_A$ ’, since this emphasises that form-restricted pluralism itself is neutral with regards to monism and pluralism at the atomic level.)

Asides from satisfying the relevant constraints, here is the central attraction of the view. *Any* inflationist will concede that all and only the true first-order negations are  $T_{\neg 1}$ ; all and only the true first-order conjunctions are  $T_{\& 1}$ ; all and only the true first-order disjunctions are  $T_{\vee 1}$ ; and so on, for *all* of the form-restricted truth properties I’ve postulated. To this extent, the proposal is utterly non-revisionary; that the relevant truths *have* these properties is simply common ground. But while any monist will concede that a true first-order conjunction is the conjunction of two truthbearers that are  $T_A$ , what she denies is that this is what its truth consists in. Instead, she postulates that its truth consists in the same thing that the truth of the components consists in:  $T_A$ . (Furthermore, if the explanatory reading of the extensionality constraints is right, then they are even committed to thinking that a first-order conjunction is  $T_A$  *because* it is  $T_{\& 1}$ .) Ontologically speaking, then, the monist is committed to everything the form-restricted pluralist is committed to. So, despite the “infinite proliferation” of truth properties, my theory is *not* more ontologically committed than the monist’s. On the contrary, what the monist does is add a *further* metaphysical assumption to the form-restricted view: that all the truths are *also*  $T_A$ , and that this is what their truth consists in. This puts the monist on the dialectical backfoot: we ought to resist adding the monist’s further assumption until we are

given good reason to add it.<sup>23</sup> My contention in this chapter is that neither logic nor the metaphysics of complex truth gives us such a reason.

Given the infinite proliferation of truth properties, we cannot write down the comprehensive metaphysics of truth. But what we have is a *recipe* for giving the metaphysics of truth for *any* truth-functional complex. That is the sense in which the metaphysics is comprehensive.

We hereby have a pluralist metaphysics of truth for complexes that is monist- and pluralist-friendly at the atomic level. If this is satisfactory, the Problems of “Mixed” Inferences and Compounds are dissolved. I’ve made the case that, other things being equal, the suggestion is to be preferred to the monist alternative. To make the case that other things *are* equal, I respond to objections in the next subsection.

### 6.2.2 *Objections and Replies*

My view is that we should see principles like “an argument is valid iff: necessarily, if the premises are true, then the conclusion is true” and “a conjunction is true iff both its conjuncts are true” as enforcing *structural* constraints on our metaphysics – constraints on the *relations between* the properties relevant for the relevant truthbearers – that *may* be satisfied by an identity (though merely postulating an identity does not guarantee satisfaction), but need not be. And I have provided a metaphysics that satisfies these constraints, as I have construed them, without postulating an identity. Is there any reason *as far as logic and logical form are concerned* to add the monist’s further assumption to our metaphysics?

*Couldn’t we avoid all the messy complexity and the infinite proliferation by simply offering a recursive account of the nature of truth for complexes, a la Tarski? And wouldn’t this allow us to define an unrestricted truth property?* I don’t think so. When it comes to *metaphysics*, I worry that building recursivity into the *nature* of a property renders it self-grounding in a problematic sense. In any case, consider how someone who buys a recursive account of the nature of truth would *explain* the truth of a higher-order complex, like ‘ $((p \ \& \ q) \ \& \ (r \ \& \ s)) \ \& \ ((t \ \& \ u) \ \& \ (v \ \& \ w))$ ’. They would say that it is true because its conjuncts are true; but when pressed to explain the truth of the conjuncts, they would say that they are true because both of *their* conjuncts are true; and when pressed to explain the truth of them, they would say that they are true because of both of *their* conjuncts are true. That is, *the explanation of the truth of the complex would be identical to the form-restricted pluralist’s*. So, in what sense does the recursive definition

---

<sup>23</sup> Though it will not be my topic until Chapter 8, I will like to point out here that this further postulation saddles us with the liar paradox: the sentence that says of itself that it lacks this unrestricted truth property. To anticipate the argument of that chapter, we avoid paradox if we do not postulate an unrestricted truth property, for I show that any liar sentence that says of itself that it lacks one or other of the form-restricted truth properties is non-paradoxical. In what follows, I will be arguing that we don’t *need* an unrestricted truth property; but part of what is driving me is the conviction that we don’t *want* an unrestricted property, on pain of paradox.

give us any kind of *metaphysical* unity? I don't think it does. The recursive definitions are useful as a kind of shorthand, but add nothing to our metaphysics of the relevant properties.

*What about other complexes?* We have not said what truth consists in for *quantificational* truthbearers yet. But I imagine the reader can guess what I am going to suggest here. Truth for quantificational truthbearers is generally understood in terms of satisfaction, and there is both a popular and an unpopular way of doing this. The unpopular way is to understand satisfaction in terms of truth, in which case satisfaction of quantifiers of different logical complexity will be understood in terms of different complex truth properties. The popular way is to understand satisfaction recursively. Once again, I think the *metaphysics* of satisfaction is hereby pluralistic in precisely the same way that the metaphysics of complex truth is. Unsurprisingly, then, I endorse a form-restricted *satisfaction pluralism* given either approach.<sup>24</sup>

There are non-truth-functional complexes, and any comprehensive metaphysics of truth must tell us what their truth consists in. For instance, the truth of the components of the following is necessary but insufficient for their truth:

- (3) Ajax left the room because there was a loud noise.
- (4) Eating Ajax is wrong because he is a household pet.

I will not take any stance on what truth consists in for (3) and (4) here. However, two points need to be made. First, I hope it is clear by now that it would be absurd to suggest that there is any special problem *for the pluralist* when it comes to such compounds. It is the *monist* who is ideologically committed to giving precisely the same answer here as she gives everywhere else. The pluralist can go in for whatever theory of truth seems most plausible. For instance, one can imagine an argument that (3), as a descriptive assertion, is apt for correspondence ('because' here perhaps denoting a causal relation), while (4), as a moral evaluation, is apt for weak super-stability; or perhaps some other theory of truth again will be appropriate. Second, such truthbearers can themselves be components of truth-functionally complex truthbearers. For these purposes, they should be treated as atomics, and the relevant truth properties for the complexes of which they are components will thus be grounded in their relevant truth properties, whatever they may be.

*What's validity, then? What's a conjunction, then?* Both of the principles mentioned at the start of this subsection mention truth more than once. But my theory is that the nature of truth varies for those truthbearers mentioned each time (the premises and the conclusion on the one hand; the conjunction and its conjuncts on the other). Are these principles not equivocating? How can we

---

<sup>24</sup> Shapiro suggests a domain-restricted satisfaction pluralism as an amendment to Lynch's functionalism – see <http://ndpr.ndu.edu/news/truth-as-one-and-many/>.

make sense of them? There are two challenges here. One is to explain the semantic function of ‘is true’, given the underlying metaphysical diversity. I delay this until Chapter 8. The substantive question for our purposes here is when we read these principles as telling us the *nature* of validity or conjunction. If the underlying metaphysics of truth is disunified, are the phenomena of validity and conjunction disunified?

If one does wish to understand the nature of validity and conjunction in terms of these principles, my suggestion is simply that we read them as articulating *structural* relations on the relevant truth properties, and understand validity and conjunction in these terms. A valid inference is one such it satisfies SVC\*. A conjunction is a compound such that its truth depends on the truth of its conjuncts in the way articulated. These structural similarities are explanatorily potent! One might, for example, explain why one should only believe a conjunction when one should believe both of its conjuncts by citing it.

One might instead understand a conjunction as being any truthbearer whose main connective is <and>, where this concept is explained in terms of its inferential role (which can be articulated while mentioning neither truth nor validity). This has the virtue of *explaining* the conjunctive structural constraint on  $T_{\&1}$ : the relevant truth property for a conjunction is one it has just in case its components have their relevant truth properties because one can infer a conjunction from both its conjuncts, and can infer each of its conjuncts from a conjunction. This explains the principle in a way that is perfectly consistent with the disunity of truth. I do not see any block to understanding validity or logical operations as unified phenomena resulting from the disunity of truth. This should be no surprise, since these are not typically pressed as challenges against the *deflationist*, whose disunity is more radical than mine.

My challenge to the objector, then, is to articulate what exactly is *missing* from our conception of logic if we do not think of it as concerned with tracking the “movement” of a single property as it travels hither and thither across inferences, and instead think of it as modelling the structural relations between truth properties in order to obtain general theories as to the nature of good inference. Granted, the logician typically assigns single *values* – canonically ‘true’, ‘T’, or ‘1’; and ‘false’, ‘F’, or ‘0’ – in the course of her theorising. But by the form-restricted pluralist’s lights, the use of a single value is an artefact of the model: the purpose is to abstract away from the particular features of the properties to model their structural relations. Note especially that the structural relations between the truth properties is not peculiar to the form-restricted view, but is common ground; the monist merely postulates an additional metaphysical unity to the properties. So this view of logic can be – as conceptions of logic should be – neutral as to the underlying metaphysics of truth.<sup>25</sup>

---

<sup>25</sup> A similar point should stand in, say, semantics, where we use ‘true’ and ‘false’ as the semantic values of sentences or propositions. Compare Yu (2017).

*Unified explanations and norms?* It may seem odd that an inflationist like me, who has been arguing for the explanatory and normative import of truth, should be ready to defend the radical disunity of the nature of truth. To put the point another way: haven't I now frustrated the whole point of being an inflationist by postulating that truth has a radically disunified nature after all? But I just do not see how this is supposed to follow.<sup>26</sup> The form-restricted disunity of truth is simply consistent with the explanatory and normative roles that I think correspondence and weak super-stability play.

Consider again the explanatory reading of the extensionality constraints on truth for logical complexes. Given this, every inflationist should be agreed that the truth of, say, a conjunction, whatever it consists in, is grounded in the truth of its conjuncts, whatever that consists in. Given *this*, the explanatory and normative import of this property should be grounded in the components' truth properties. If, for instance, we explain the practical success of an agent's action in part because of the truth of her belief in a conjunction, this conjunction is true *because* its conjuncts are true, and thus the explanation is ultimately grounded in the correspondence between the conjuncts and the world. Similarly, if we are to explain why someone ought to believe a moral conjunction, since this moral conjunction is true *because* its conjuncts are true, the explanation will be grounded in each of the conjuncts being weakly super-stable. The form-restricted truth properties I endorse are ultimately grounded in correspondence and weak super-stability, which are explanatorily and normatively fundamental. Their import is derivative.

Indeed, *even the monist* will say this. The monist might think that, for any  $p$ , one ought to believe that  $p$  iff  $p$  corresponds (say), but if  $p$  is a first-order conjunction, she will think it corresponds *because* its conjuncts correspond; and thus that one ought to believe that  $p$  because its conjuncts correspond. The fact that the complex itself corresponds looks explanatorily and normatively *redundant*. All the work will be done by the form-restricted properties, the existence of which is (to reiterate) common ground.<sup>27</sup>

## Conclusion

I have argued that the Problems of "Mixed" Inferences and Compounds are just particular instances of perfectly general, structural constraints on our metaphysics of truth, which the monist does not automatically satisfy simply by postulating that truth always and everywhere consists in the

---

<sup>26</sup> One can, of course, be a form-restricted pluralist and a moderate pluralist, by thinking that the form-restricted properties too are versions of  $T_U$ . My argument here shows that  $T_U$  is redundant as far as the Problems of "Mixed" Inferences and Compounds are concerned. Chapter 8 gives us good reason to reject it.

<sup>27</sup> I am thus even tempted to put forward the rather more extreme suggestion that, *even if* it should turn that a complex *does* have the monist's favoured truth property, she should not say that this is what the truth of the complex consists in. As far as the truth of complex is concerned, it looks *superfluous* (or, perhaps more politely, *supererogatory*). The form-restricted properties are sufficient. Again, the ultimate justification for this comes from resolving the liar paradox; see Chapter 8.

same thing. I have then proposed a form-restricted pluralism about the nature of truth that satisfies these constraints, which is neutral between monism and pluralism at the atomic level, and which I have suggested all inflationists should be sympathetic towards. And I have defended the resultant view from the most obvious objections. I cannot claim to have been comprehensive in this defence, but I leave it as a standing challenge to the would-be objector to articulate some shortcoming this form-restricted pluralism has with regards to logic or logical form.

My theory is thus that truth is substantive but radically disunified, albeit in a sense I think is ultimately quite familiar and considerably less shocking than it might at first appear. Thus far, I have been arguing that we *can* do without an unrestricted truth property. In the final chapter, I'll argue that we *ought* to do without one. We must first, however, deal with a “real” mixing problem: the Problem of Mixed Atomics.



---

# CHAPTER 7

## DISENTANGLING TRUTHS

---

‘It is lovers of thick concepts who deny that there are two things we are doing. They say that we are doing one thing, describing something as courageous, for example. Expressivists recognize that there are two activities going on and that we manage to do both things at once.’

Blackburn (2013b: 123).

The Problems of “Mixed” Inferences and Compounds concentrate on cases where truthbearers the pluralist thinks are apt for different properties are “mixed” together. The Problem of Mixed Atomics concerns what to say about *atomic* truthbearers that seem to “mix” content from different domains.<sup>1</sup> The prime examples I’ll focus on here involve so-called “thick” ethical predicates or concepts.<sup>2</sup> In this short chapter, I show how we can deploy analogous moves to those found in the literature on thick ethical predicates or concepts to solve the Problem as it arises for my metasemantically-driven pluralism.

### 7.1 The Problem of Mixed Atomics

#### 7.1.1 Domain Individuation

There is no such thing as *the* Problem of Mixed Atomics, as such. Distinguish the following two matters. The first concerns domain individuation: how to sort truthbearers into those that are apt for  $T_1$  and those that are apt for  $T_2$ . The second concerns what to say about those atomic truthbearers that, once this sorting is done, are either sorted into more than one domain or no domains.<sup>3</sup> The latter is the Problem of Mixed Atomics. How serious or widespread the Problem is – what exactly it amounts to – depends on one’s account of domain individuation, as well as the nature of the truth properties involved.

---

<sup>1</sup> See David (2009: §8.2; 2013), Lynch (2009: 78-82) and especially Wyatt (2013) and Stewart-Wallace (2016).

<sup>2</sup> This is a research topic in its own right with a substantial literature: see Väyrynen (2016) for an excellent overview, Väyrynen (2013) for a book-length treatment, and also the papers collected in Kirchin (2013).

<sup>3</sup> It’s a little odd to say that atomics in *no* domain are “mixed”, but whether or not an atomic counts as being in no domains or more than one typically turns on a terminological choice concerning domain individuation. See, e.g., fn.9 below. For ease, I’m also talking as though we’ve settled the nature of truth for all different discourses.

Focus for the time being on sentences. Remarkably, many of the examples that have been the focus of discussion in the literature do not look like mixed atomics at all:

- (MA1) Killing is morally wrong.
- (MA2) Immoral acts happen in space-time.<sup>4</sup>
- (MA3) Charlie is delicious.<sup>5</sup>
- (MA4) The number 17 is beautiful.<sup>6</sup>
- (MA5) This crystal is beautiful.<sup>7</sup>

For one thing, (MA1) and (MA2) are plausibly universal generalisations (and, for a Russellian, so is a definite description like (MA4)).<sup>8</sup> But in any case, (MA1) is a paradigmatic moral sentence, while (MA3), (MA4), and (MA5) are paradigmatic aesthetic sentences. If anyone thinks that the nature of truth varies between moral or aesthetic truths on the one hand, and those concerning physics or medium-sized dry goods on the other, it is almost certainly cases like (MA1), (MA3), (MA4), and (MA5) that they have in mind. Why, then, think that these cases are “mixed” at all?

The best explanation is that the theorists have been working with a particular understanding of domain individuation that renders them so; this is explicit in Lynch (2009: 78-82) and especially Wyatt (2013: S230). As emphasised in Chapter 2, ontologically-driven motivations for pluralism dominate the literature at present. For ontologically-driven pluralists, it is natural to want to individuate domains (how else?) ontologically. To this end, Wyatt assumes that there are different kinds of individuals, properties, and relations: e.g., mathematical individuals, properties, and relations; ethical individuals, properties, and relations; aesthetic individuals, properties, and relations; scientific individuals, properties, and relations. He calls these “topics”. Language and thought are then divided into parts that are about different topics, presumably in virtue of referring to, denoting, or ascribing the relevant individuals, properties, and relations. A sentence or proposition is then in a particular *domain* just in case it concerns only individuals, properties, and relations of a particular *topic*.<sup>9</sup> On this understanding, (MA1)-(MA5) are not members of one particular domain: (MA3), for instance, because (presumably) while ‘Charlie’ refers to a physical object (a beetroot), ‘is delicious’ ascribes an aesthetic property: the individual and the property are parts of different topics.

This view is certainly at least implicit in all focused discussion of the Problem so far, and Wyatt has done a service in making it explicit, but it is rather surprising. For it transpires that only a

---

<sup>4</sup> (MA1) and (MA2) are from David (2009: §8.2).

<sup>5</sup> Wyatt (2013: S233) – ‘Charlie’ is the name of a beetroot.

<sup>6</sup> Lynch (2009: 79).

<sup>7</sup> David (2013: 50, fn.9).

<sup>8</sup> (MA1) might be a generic.

<sup>9</sup> The word ‘only’ here means that the “mixed” cases belong to *no* domain; without it, they belong to more than one. The difference doesn’t matter for our purposes.

very restricted set of truthbearers get to fall within, e.g., the aesthetic or moral domains: it is only moral or aesthetic evaluations of other moral or aesthetic properties or relations that get to be part of the moral or aesthetic domain (it is hard to think of what an aesthetic or moral *individual* might be). Wyatt (2013: S234) suggests ‘Deliciousness is awesome’; a moral example might be ‘Evil is bad’. This is remarkable. It excludes the vast majority of cases with which pluralists are concerned: ‘Eating meat is wrong’, ‘*Deadwing* is beautiful’, ‘Paul Foot is funny’, etc. are *all* ruled as mixed atomics.<sup>10</sup> It turns out the domain-restricted pluralist’s theory is not so much that truthbearers in different domains are apt for different properties, but that different sets of mixed atomics are apt for different properties. The principle issue being that the pluralist then needs to tell us *which* mixed atomics are apt for which truth properties, which is just the question of domain-individuation all over again! And once we have this additional account, which *does* sort paradigmatic cases like (MA1)-(MA5) into their right domains, it’s not obvious what work the one that only sorts cases like ‘Deliciousness is awesome’ and ‘Evil is bad’ is going to do. No, the ontologically-driven pluralist is better off rejecting this account and finding another.<sup>11</sup>

In his forthcoming book, Edwards (fc) suggests that the individuating work is done, for atomics of the form ‘a is F’ where ‘a’ is a name and ‘is F’ a predicate, by the predicate; and, specifically, the kind of property it ascribes.<sup>12</sup> This is essentially an adaptation of Wyatt’s proposal, which denies a role for the name in domain individuation. This strikes me as a far more plausible criterion of individuation, and assuming that it can be generalised to sentences using definite descriptions and complex demonstratives instead of names, it immediately dissolves any impression that any of (MA3)-(MA5) are mixed.<sup>13</sup> If someone could formulate a convincing ontologically-driven argument for truth pluralism, I think this would be the account of domain individuation to use.<sup>14</sup>

But, unsurprisingly, I do not like it for the moral domain: I do not like the suggestion that moral properties are of a different kind to physical properties – this has the unpleasant whiff of non-naturalism. In any case, the expressivist allows that moral properties are natural properties (§4.2.3). The metasemantically-driven pluralist will instead individuate truthbearers (how else?) metasemantically. It depends on how we explain the meaning of the relevant sentences – see Chapter 4. Nonetheless, this will align in the relevant cases with Edwards’s proposal, since it is moral *predicates*

---

<sup>10</sup> This view must lie behind Stewart-Wallace’s (2016: 365) idea that a Millian theory of names might give rise to mixed atomics: ‘Names, like ‘Tom’, if their semantic value is just their referent, i.e. Tom, presumably take Tom *warts and all*, with his moral, mathematical and physical characteristics.’ In Wyatt’s terms, the thought seems to be that ‘Tom’ hereby concerns a variety of topics.

<sup>11</sup> As an educated guess, I’d suggest that the prevalence of this rather unhelpful account of domain individuation comes in part from focusing on the mathematical case, since mathematical truthbearers are, in many cases at least, *exclusively* about mathematical individuals, properties, and relations.

<sup>12</sup> In the draft I have seen (cited here with permission), Edwards seems to see his account of domain individuation as linguistically-driven, claiming that different sets of predicates play different “functional roles”. But it turns out these “functional roles” pretty much just involve denoting different kinds of properties; so the difference is ontological after all.

<sup>13</sup> (MA1) and (MA2) both involve more than one predicate; but, as mentioned, they are plausibly non-atomic.

<sup>14</sup> This only works for *sentences*; at the level of proposition or belief, we’d presumably need individuation to be done by the concepts voiced by predicates.

that are used to express desire-like states, where descriptive predicates are used to express (maximal) beliefs.

### 7.1.2 Mixed Predicates

To find a genuine mixed atomic sentence, then, what we need to find is a mixed *predicate*.<sup>15</sup> And here the literature provides, for there is already much discussion of so-called “thick” ethical predicates (or concepts), which seem to “mix” both descriptive and evaluative content.<sup>16</sup> Predicates like ‘courageous’, ‘just’, ‘lewd’, ‘chaste’, ‘rude’, and so on. To call something courageous, just, or chaste is to evaluate it as good in a certain way; to call something lewd or rude is to evaluate it as bad in a certain way. But thick ethical terms stand apart from “thin” or “purely” evaluative terms, like ‘good’, ‘bad’, ‘right’, and ‘wrong’, in their *specificity*. Not just any good act is a courageous one: fetching some shopping for an elderly relative is certainly a good thing to do, but it is not (in the usual run of things, anyway) a *courageous* thing to do. Similarly, leaving an injured animal to suffer painfully to a slow death is typically a bad thing to do, but it is not *rude* (let alone *lewd!*). On the face of it, then, it looks as though such predicates incorporate *descriptive* information, such that saying that an action was courageous is not just to *evaluate* it as good, but also to *describe* it as having certain features (perhaps in virtue of which it is good).

The question for the truth pluralist is what property an atomic sentence like

(C1)            x was courageous.

is apt for.<sup>17</sup> Note, however, that it would be too quick to charge: “It’s neither in such-and-such domain, and so apt for T<sub>1</sub>, nor in such-and-such domain, and so apt for T<sub>2</sub>; therefore, it’s neither apt for T<sub>1</sub> nor T<sub>2</sub>!” The pluralist is not committed to truth properties only being relevant for one domain each. What the pluralist will want to say about these cases will depend on why, exactly, she is a pluralist. Someone motivated by a proto-scope problem (§2.1) might think that the evaluative element of (C1) renders correspondence “implausible”, but not think that the descriptive element renders superwarrant “implausible”. Indeed, both Lynch’s concordance (§1.2.3) and my weak super-stability (§5.4) allow correspondence of *other* truthbearers to play a role in determining which

---

<sup>15</sup> That there are two related issues here – domain-individuation and mixed atomics – has perhaps been insufficiently clear. I think this is because the account of domain individuation made explicit by Wyatt has been taken for granted. On mine and Edwards’s accounts, there is no Problem of Mixed Atomics besides the Problem of Mixed Predicates; for Wyatt and Lynch, this is an instance of a more general problem.

<sup>16</sup> I am not strictly committed to *all* “descriptive” truthbearers being apt for correspondence (e.g., those outside the scope of the Success Argument), nor *all* “evaluative” truthbearers being apt for weak super-stability (e.g., non-moral normative thought and talk) – although I am at present inclined this way. However, I am certainly committed to *some* descriptive truthbearer being so, and *some* evaluative truthbearer being so, so I follow the literature here in talking about descriptive and evaluative content generally.

<sup>17</sup> The specific focus on predicates and sentences here is quite intentional.

truthbearers concord/are weakly super-stable. Perhaps this does sufficient justice to the descriptive element. One can imagine a principled argument to the effect that *any* evaluative element is sufficient to render concordance or weak super-stability relevant for a truthbearer.

Again, then, there is no *automatic* problem here. We need an explanation of what the truth of (C1) consists in; and it is up to the pluralist to provide one that is consistent with her pluralism.

I am tempted by the above line of thought – that the truth of (C1) consists in weak super-stability – but prefer the approach articulated below for a couple of related reasons. The first is that, if a thick ethical judgement does mix descriptive and evaluative content (on which more shortly), the two theoretical roles for truth that I have argued are played by different properties become *entangled*. It will be easiest to illustrate this using a simple example. Suppose we define the word ‘badwork’ thus:

*Badwork*

$\forall x(x \text{ is badwork} \leftrightarrow (x \text{ is time-consuming} \ \& \ x \text{ is wrong}))$ .

Assuming that judgements of time-consumingness fall within the scope of the Success Argument, while judgements of wrongness are moral. Particular instances of ‘x is badwork’ will then be mixed atomic sentences, and may be true. Suppose, then, that I judge that some activity is badwork, and that this is true. Actions that result from this belief may be (non-coincidentally) successful because the belief is true. Suppose, for instance, that I distract someone for long enough to break into the safe by getting them to do some badwork. If I am right that correspondence explains practical success (§4.1), then I am committed to explaining this success via correspondence.<sup>18</sup> But it cannot be mere correspondence that norms the belief that something is badwork, since it involves evaluative judgement, and evaluative judgements I have argued (Chapter 5) are normed by weak super-stability.<sup>19</sup> Now, this example is contrived, but if thick ethical predicates do mix (relevant) descriptive and evaluative content, then there is no reason that these roles should not become entangled in an analogous manner. This puts pressure on me to disentangle these roles by disentangling the descriptive and evaluative content of the judgement: to anticipate somewhat, the view will be that it is the correspondence of the descriptive elements that explains practical success, while weak super-stability norms the evaluative elements. The truth of the complex “mixture” of both will then be understood in terms of a complex truth property “built” out of both

---

<sup>18</sup> Compare Blackburn (2013b: 129) on explanations involving pig-headedness. Blackburn’s view is that there *are* “patterns” of success of the kind he thinks are distinctive of (what I call) maximal beliefs when it comes to thick ethical judgements (Blackburn’s “loaded” terms), which there are not when it comes to (thin) moral judgements. Nonetheless, in these cases “[p]erhaps the [evaluative] load does none of the explanatory work.” I take Blackburn’s view to be highly congenial to my own. See also §4.2.3, fn.23.

<sup>19</sup> Perhaps there is some way of explaining success in terms of weak super-stability here given the role correspondence plays in determining the weakly super-stable beliefs. I can’t see this at present, but if so this is only good news for me.

correspondence and weak super-stability, in the manner described in the last chapter (§6.2.1). The second reason is that, qua expressivist, I am *already* committed to being able to disentangle the descriptive and evaluative content of mixed atomics like (C1). I explain why in the next section.

## 7.2 Disentangling Truths

### 7.2.1 Expressivist Approaches to Thick Predicates

Thick ethical predicates are meant to provide a standing challenge to expressivists: if (C1) has both descriptive and evaluative content, how are we to explain its meaning? Does it express a desire-like state or a maximal belief? Qua evaluation, it seems as though the expressivist is committed to explaining its meaning in terms of some desire-like state that it expresses. But this cannot be simple approval, for then (C1) would mean the same as ‘x is good’.<sup>20</sup>

A natural response strategy for the expressivist is to maintain that predications of thick ethical terms like (C1) can be *analysed* into a non-evaluative description and a thin evaluation. A simple proposal on these lines might be:

(C2)            x has such-and-such features and x is good (because of it).

Where ‘such-and-such features’ constitutes a non-evaluative description. We will be precise about what purpose such analyses are meant to serve below. But first we ought to note that the prospects for such a simplistic analysis have been criticised on the grounds that thick ethical terms are descriptively “shapeless”; that is, roughly, the thesis that there is no substantive non-evaluative feature that is shared by all and only the courageous actions. (This claim is often evidenced by hypothetical claims of competence: that one could not master the extension of a thick ethical term without participating in an evaluative practice.)<sup>21</sup> The thought is that the extension of a thick ethical term thus cannot be fixed by some non-evaluative description (“has such-and-such features”) that is part of its content. Some critics of expressivism have argued on this basis that thick ethical terms must track distinctively ethical features of the world, in precisely the representational way that the expressivist denies.<sup>22</sup>

---

<sup>20</sup> Using approval here as a stand-in for whatever state the expressivist says is expressed by ‘is good’.

<sup>21</sup> E.g., Williams (1985) drawing on McDowell (1981).

<sup>22</sup> If the critics here are right, then this may be good news for the ontologically-driven pluralist qua Problem of Mixed Atomics. While some take the upshot to be that thick predicates “fuse” their non-evaluative and evaluative content such that they can’t be disentangled (Williams 1985, Kirchin 2010, Harcourt & Thomas 2013), one may think, with Roberts (2013), that the right moral would be that thick predicates don’t really “mix” descriptive and evaluative content at all – the normativity is primarily in *the world*, not the content. (See Eklund (2013) for a compelling (and expressivist-friendly!) worry about locating evaluation here.) But if thick ethical predicates are evaluative in virtue of ascribing distinctively normative properties, then they’re not really mixed – indeed, this sounds like *exactly* the type of thing ontologically-driven pluralists say. In the grander scheme of things, however, it might be bad news, for this road typically leads either from or to moral realism,

Now, it is beyond our purview here to settle this debate. (Needless to say, if the contention of the final line of the last paragraph is right, then one of the underlying assumptions of this thesis is wrong, and I have bigger problems than mixed atomics to worry about.) What is important for present purposes is that Elstein & Hurka (2009) have shown that it is possible for the expressivist to allow evaluation a role in determining the extension of a thick ethical predicate. To this end, they offer two “reduction plans” for analysing thick ethical terms into non-evaluative description and thin evaluation illustrated by the following proposed analyses of ‘distributively just’ (D<sub>EH</sub>) and ‘courageous’ (C<sub>EH</sub>) respectively:<sup>23</sup>

(D<sub>EH</sub>)            x is good, and there are properties X, Y, and Z (not specified) that distributions have as distributions, or in virtue of their distributive shape, such that x has X, Y, and Z, and X, Y, and Z make any distribution that has them good.

(C<sub>EH</sub>)            x is good, and x involves an agent’s accepting harm or the risk of harm for himself for the sake of goods greater than the evil of that harm, where this property makes any act that has it good.<sup>24</sup>

In (D<sub>EH</sub>), the unspecified properties X, Y, and Z are identified as being properties that things have in virtue of being distributions, which restricts the application of ‘distributively just’ to just those things that have certain non-evaluative features, i.e., are distributions. But determining its extension requires evaluative judgement of *which* features make distributions good. Thus, on this analysis, evaluation plays a role in determining the extension of ‘distributively just’. (C<sub>EH</sub>) is similar, but also *embeds* a thin evaluation: determining which things are harms and which goods and in what proportion such that an agent is acting for the sake of goods greater than the harms risked. In this way, the expressivist can concede that evaluation drives the extension of thick terms while still maintaining that they can be analysed in terms of non-evaluative description and thin evaluations. We will return to this strategy shortly.

First, however, we need to briefly note two alternative expressivist strategies for dealing with thick ethical terms, only to set them aside. The strategy is to *deny* that thick ethical terms mix descriptive and evaluative content as part of their semantic content. There are two ways of doing so.

---

so whence the block to correspondence? (This is the same kind of worry again for ontologically-driven pluralists as I raised in Chapter 2.)

<sup>23</sup> It does not matter for our purposes whether or not these are the right analyses of these terms. Indeed, it does not matter for our purposes if any such analysis *can* be given in any *actual* language, but only that it *could* be given.

<sup>24</sup> Elstein & Hurka (2009: 522, 527).

The first – endorsed by Blackburn (1992, 1998a, 2013b) and (for independent reasons) Väyrynen (2013) – is to hold that thick ethical terms are not evaluative as part of their semantic content or truth conditions, but are merely *used* to convey an evaluation, i.e., that they typically convey evaluative content as a result of some pragmatic mechanism, such as conventional implicature or presupposition. For instance, it is common to hold that ‘Ed is strong but smart’ has the same truth conditions as ‘Ed is strong and smart’ (namely, Ed’s exemplifying both of these properties), but conventionally implies by the use of the word ‘but’ that there is a contrast between strength and intelligence, such that the latter is unexpected given the former. Similarly, perhaps to call something “lewd” is just to say that it is an overt sexual display (over a certain threshold of overtness or sexualness, perhaps), but the use of the word ‘lewd’ conventionally implies that such displays are bad in a certain way.

The second – endorsed by Gibbard (1992) – is to maintain that thick ethical terms, like thin ethical terms, are purely evaluative: that is, each is to be explained in terms of some desire-like state that it expresses. As mentioned above, a negatively evaluative term like ‘lewd’ cannot express disapproval, for then there would be no difference in meaning between ‘x is lewd’ and ‘x is bad’. This proposal thus requires that ‘lewd’ expresses a *distinct* desire-like state, which Gibbard labels ‘L-censoriousness’. This attitude is only warranted (if it is warranted at all) towards overt sexual displays, which explains its increased specificity. Generalising this story thus requires that each thick ethical term expresses a different desire-like state.<sup>25</sup> But there is no reason, in principle, that the expressivist should not combine these approaches in a piecemeal fashion. Perhaps certain thick ethical terms are evaluative as a matter of pragmatics, some express distinct desire-like states, and yet others permit of analysis into non-evaluative description and thin evaluation.

As indicated, it is beyond our purview to settle what the right account of thick ethical terms is, but to sustain the contention that they present some insuperable difficulty for the expressivist requires arguing that these strategies between them are incapable of explaining them (and, once again, failure is a problem for expressivism in general, not the resultant alethic pluralism). We shall therefore assume that one of the approaches outlined above is correct for each thick ethical term.

As far as the *truth* of thick ethical predications is concerned, *which* answer is right might look to be of vital import. After all, if either Blackburn or Gibbard is right, then thick ethical predications do not constitute mixed atomics after all: if Blackburn is right, then they are merely descriptive, and hence (probably) apt for correspondence; and if Gibbard is right, then they are merely evaluative, and hence apt for weak super-stability.

However, it is actually quite inconsequential for our purposes which account is right. For whichever semantic theory of ‘courageous’ is *actually* correct, there *could* have been a word in the English language – and we can introduce one right now, ‘courageous<sub>EH</sub>’ – which permits of the

---

<sup>25</sup> The exegetical issue of whether or not this is exactly Gibbard’s view does not matter here; I am interested in the view as representative of a certain strategy.



analysis ( $C_{EH}$ ). The sentence ‘ $x$  is courageous $_{EH}$ ’ would then constitute a mixed atomic, which is the problem case. Therefore, even if there aren’t, in fact, any mixed atomic sentences in the English language, we can introduce them; and both the expressivist and the pluralist will need a story concerning the meaning and truth of such sentences respectively. Moving forward, then, I will assume that ( $C_{EH}$ ), or something like it, is the right analysis of the English word, ‘courageous’.

I’ve been focusing on thick ethical predicates here because it’s the relevant case that has received most attention in the literature, and because it is most pertinent to my own metasemantically-driven pluralism, but this latter point allows us to generalise the discussion. For it is easy to *build* a mixed atomic by definition, in the manner of ‘badwork’ above. Take any predicate ‘is  $F$ ’ that the pluralist thinks is in one domain and another ‘is  $G$ ’ from another, and we can define a predicate ‘is  $H$ ’ thus: for any  $x$ ,  $x$  is  $H$  iff ( $x$  is  $F$  and  $x$  is  $G$ ). And we can do this using any number of predicates and fanciful logical operations on the right-hand side of the definitional biconditional. The resultant predicate will allow us to formulate mixed atomic sentences. So, no pluralist can avoid the Problem of Mixed Atomics, even if her account of domain individuation sorts all sentences of the English language (or whichever) *as is* into disjoint domains.

### 7.2.2 Disentangling Truths

We’re assuming, then, that we can disentangle the descriptive and evaluative content of a thick ethical predication like (C1) via an analysis such as ( $C_{EH}$ ). This provides an opening for the pluralist to reduce the Problem of Mixed Atomics, of which (C1) is an instance, to the Problem of Mixed Compounds, of which ( $C_{EH}$ ) is an instance. The two are *ex hypothesi* synonymous; each is true just in case the other is. What the pluralist needs is an argument that the analysis has a certain kind of *alethic priority* over the atomic.

As discussed earlier (§1.1.2), it is quite usual to maintain that certain *kinds* of truthbearer are non-primary, in the sense that their truth is to be understood in terms of some other kind of truthbearer. As sentences, how exactly we cash out the alethic priority of ( $C_{EH}$ ) over (C1) will depend on what we take the primary truthbearers to be. If sentences are non-primary, then strictly neither ( $C_{EH}$ ) nor (C1) is a primary truthbearer. Since they will, qua synonyms, be associated with the *same* primary truthbearer, however, what the pluralist will want to say is that ( $C_{EH}$ ) is revelatory of the logical form of the relevant primary truthbearer in a way that (C1) is not. In this way, we can “read off” the relevant truth property for the primary truthbearer from the logical form (and subject matter) of ( $C_{EH}$ ), but not (C1).

For instance, if propositions are the primary truthbearers, then both ( $C_{EH}$ ) and (C1) voice the same proposition, but the logical form of ( $C_{EH}$ ) mimics that of the proposition they voice. This is, in fact, my preferred option. It is highly plausible that certain simple linguistic entities – words; specifically, in this case, predicates – voice complex concepts. This is most plausible for those terms

explicitly introduced via complex definition: e.g., ‘badwork’ plausibly voices the concept <time-consuming and wrong>. (The component concepts, like <time-consuming>, may also be complex, of course.) But this is not only plausible for definitions that mix different kinds of content. If I define ‘gbjag’ as ‘either a triangle or square, or both a hamster and old’, then ‘gbjag’ voices a complex concept. On my preferred way of thinking about this topic, thick ethical predicates voice complex concepts; ones composed out of “purely” descriptive and evaluative concepts. The analysis aims to reveal the complex structure of the concept so voiced. If, however, beliefs are the primary truthbearers instead of propositions, then the view is that (C<sub>EH</sub>) and (C1) both express the same complex psychological state, but that (C<sub>EH</sub>) is revelatory of its complexity, where (C1) hides it.

Note, further, that claiming that (C<sub>EH</sub>) has an alethic priority over (C1) does not commit us to thinking that sentences are not the primary truthbearers. It commits us to thinking that *some* sentences are non-primary. On this way of cashing things out, the pluralist will need to argue that the truth of some sentences is to be understood in terms of *translation* into the primary sentences. (Translation is here playing something like the role of voicing or expression in the other stories.) The core job here, then, is to give a principled story about which sentences are primary, and which are not.

The pluralist requires, then, that some sentences are revelatory of the logical form of the relevant primary truthbearer (perhaps, at the limit, in virtue of *being* the relevant primary truthbearer), while others are not. Call the former *canonical* sentences, while the latter are *non-canonical*. If sentences are not the primary truthbearers, it follows from the existence of synonymous sentences of different logical forms and form-restricted pluralism that sentences divide into those that are canonical and those that are not.<sup>26</sup> To reduce the Problem of Mixed Atomics to that of “Mixed” Compounds, the pluralist needs to maintain that (C1) is non-canonical, while (C<sub>EH</sub>) is canonical – this establishes the relevant alethic priority of (C<sub>EH</sub>) over (C1).

It is difficult to formulate a knockdown *argument* that this is so, but it is certainly very plausible within the metasemantically-driven framework developed here. (I leave open whether or not the alethic priority is similarly plausible within an ontologically-driven framework.) Consider the role that the analysis (C<sub>EH</sub>) plays for the expressivist. Elstein & Hurka’s explicit goal in providing the analyses is to explain away an apparent feature of the behaviour of thick predicates – namely, their descriptive shapelessness – in terms of non-evaluative description and thin evaluation. But that is not the sole, or perhaps even the most central, challenge posed by thick ethical terms, which concerns *how the expressivist is to explain their meaning*. Again, it cannot be explained in terms of expressing approval, and we’re assuming contra Gibbard that it doesn’t express some other positively-valenced state. Its evaluative content precludes explaining it in terms of some maximal belief expressed.

---

<sup>26</sup> Canonical sentences for some primary truthbearers may or may not actually exist in English.

The analysis kicks in here: we are to explain the meaning of the atomic sentence (C1) via explaining the meaning of the complex sentence (C<sub>EH</sub>).<sup>27</sup> This, I take it, is the core idea of the strategy: the expressivist needs an independent explanation of the meanings of thin evaluative terms and non-evaluative descriptions, *and complexes composed of them*. Given all this, the primary virtue of the analysis will be that it allows us to explain the meaning of the mixed atomic sentence parasitically on the explanation of the meaning of the complex. Of course, the italicised point here requires a solution to the Frege-Geach Problem (§4.2). But the expressivist needs one of these anyway.

So, while (C<sub>EH</sub>) and (C1) have the same meaning, (C<sub>EH</sub>) has a theoretical, metasemantic priority. The explanation of *why*, I take it, is because (C<sub>EH</sub>) is revelatory of the complexity of the psychological state expressed by both sentences. Given that the expressivist is committed to thinking that evaluative judgements are different in kind from descriptive judgements – they are desire-like states where the latter are maximal beliefs – a thick ethical *judgement*, then, must be a complex state, which involves *both* maximal belief and desire-like states. Now, exactly which state is expressed by a complex sentence like (C<sub>EH</sub>), or how the states expressed by the component sentences “combine” in the case of thick ethical judgement, is a tough question in moral psychology. But for present purposes, what matters is how this relates to the non-canonicity of (C1): for one thing, the metasemantic priority of the complex sentence renders it plausible that the simple sentence voices a complex proposition (perhaps in virtue of the predicate voicing a complex concept) and it is a consequence of expressivism that it expresses a complex psychological state. Just as important is that it allows us to disentangle the theoretical roles that I have argued are played by different properties. The thick ethical judgement involves both maximal beliefs and desire-like states; the arguments of Part II establish that is the correspondence of the former that is explanatorily relevant for practical success, while weak super-stability norms the latter. For a complex judgement somehow “composed” out of these simple judgements the relevant truth property is a complex truth property, built out of the correspondence of the descriptive elements and the weak super-stability of the evaluative elements, in just the way we should expect. (For a simple example, consider how this goes for a judgement that something is badwork.) The different elements of our theory line up neatly here.

(Moreover, consider the “truth-first” approach to the Frege-Geach Problem sketched in §4.2.4. Given form-restricted pluralism, we’ll explain the meaning of (C<sub>EH</sub>) in terms of a complex truth property. Given the metasemantic priority of (C<sub>EH</sub>) over (C1), we likewise explain the meaning of (C1) in terms of this property. It is then natural to suggest that (C<sub>EH</sub>) is canonical where (C1) is not, otherwise the relevant truth property for (C1) will come apart from the property in terms of which we explain its meaning.)

---

<sup>27</sup> The particular sentence may not, in fact, exist (see fn.23), since we might lack the words that voice the relevant concepts in the English language *as is*. If this is so, then by the expressivist’s lights, our attempts to explain the meaning of (C1) will to that extent be frustrated until we *do* have the relevant words. In present terminology, this is *precisely because* we cannot formulate a canonical sentence.

We can, for present purposes, leave to another day the development of a principled story about exactly which sentences are canonical. If propositions or beliefs are the primary truthbearers, then it will be whichever sentences have components that voice or express simple propositions (perhaps composed out of simple concepts) or simple mental states; we would then ideally need a principled story about which concepts, propositions, or beliefs are simple. But what matters here is not which sentences *are* canonical, but which sentences are *not*. There will be clear cases here. First, there are sentences that use words that combine descriptive and evaluative content; within the metasemantically-driven framework, this chimes with the expressivist's commitment to disentangling the descriptive and evaluative content of thick ethical predicates. Second, as mentioned above, there are those words that explicitly have complex definitions, like 'badwork' or 'gbjag'. Within the form-restricted framework, it is highly plausible that the truth of atomic ascriptions of such complex predicates is not to be understood in terms of one or another simple truth property, but in terms of the truth of the components of its analysis.

## Conclusion

In the last chapter, I argued that the relevant truth property for a truthbearer depends on its logical form, as well as its subject matter. In this chapter, I have argued that the relevant truth property for some *sentences* cannot be read off of their surface structure. In particular, some simple linguistic entities – words – express complex psychological states and/or voice complex concepts. These sentences are non-primary: their truth is to be understood in terms of the truth of the primary truthbearer to which they are appropriately related (via the relation of expression or voicing, or perhaps translation), which will be a complex composed of unmixed atomics. This allows us to reduce the Problem of Mixed Atomics to the Problem of "Mixed" Compounds in a principled manner consistent with the diverse theoretical roles of truth, form-restricted pluralism, and the complex truth properties of the previous chapter.

---

## CHAPTER 8

# TRUTH AS NONE AND MANY

---

‘...each part of the structure is acceptable as representing a flight of steps but the connexions are such that the picture, as a whole, is inconsistent: the steps continually descend in a clockwise direction.’

Penrose & Penrose (1958: 31).

I’ve argued that the nature of truth is radically diverse, consisting in correspondence or weak super-stability at the atomic level, and an infinite diversity of different properties grounded in these two for complex truthbearers. But the predicate ‘is true’ functions as a *fully general* device for endorsement. Explaining how it does so in the pluralist setting has been known as the Problem of Mixed Generalisations. The expressive function of the truth predicate may yet give us reason to endorse an unrestricted truth property as the referent of the predicate.

However, there is famously a Big Problem with postulating an unrestricted truth property: the “liar” paradox. A proposal that has been gaining traction recently is that the liar shows us that the truth predicate voices an inconsistent concept. This idea is in *prima facie* tension, however, with the claim that truth is an explanatorily and normatively important part of our world. In this chapter, I show how we can begin to resolve this tension. A way of avoiding the paradox that was popularised by Alfred Tarski (1933, 1944) and has recently been endorsed in the pluralist setting by Aaron Cotnoir (2013a) involves arguing (i) that different truth predicates or properties –  $T_1$ ,  $T_2$ , etc. – are relevant for different sets of truthbearers, and (ii) that the relevant truth predicate or property for any liar sentence that says of itself that it is not  $T_i$  is some  $T_j$ ,  $j \neq i$ . Such proposals are typically detached from the literature on the metaphysics of truth and come across as *ad hoc*, paradox-barring, moves (perhaps explicitly). I develop a novel such solution using the machinery that has been independently motivated in Chapters 6 and 7. Truth is thus a substantive part of a classically consistent world, even if the concept <truth> is an inconsistent concept.

### 8.1 Generalisation and the Liar

#### 8.1.1 *The Problem of Mixed Generalisations*

In the literature on truth, if there is such thing as an uncontroversial claim, it is that the English predicate (assuming contra prosententialism that it is a predicate) ‘is true’ plays an expressive

role in our language, primarily as a device for endorsement (§1.2.1). It is striking that this role is *fully general*. If I say ‘Whatever you just said was true’, then I endorse *whatever it is* you just said, regardless of subject matter or logical form (providing it was of the right grammatical form, etc.). Similarly, if I say ‘Everything you said yesterday was true’, I endorse *everything* you said yesterday, regardless of subject matter and logical form. This is in *prima facie* tension with the claim that the nature of truth varies depending on the subject matter and logical form of the truthbearer, for what property is ascribed by the predicate? Naturally if we say that it is one or the other of the restricted truth properties, then this fails to explain the truth predicate’s *fully general* expressive capacity. This has been called the Problem of Mixed Generalisations, since the challenge becomes especially acute in the cases where the truth predicate is used to *simultaneously* endorse truthbearers that are apt for different truth properties.<sup>1</sup>

This is first and foremost an argument for the truth predicate expressing an unrestricted truth *concept*. Our mastery of this concept, for instance, might be explained by our mastery of the “introduction” and “elimination” rules, (T-In) and (T-Out):

(T-In)	$p$ . _____ ‘p’ is true.	(T-Out)	‘p’ is true. _____ $p$ .
--------	-----------------------------	---------	-----------------------------

Where ‘p’ is a schematic marker for declarative sentences and single-quotes function as an operator for taking a sentence to the relevant primary truthbearer. As I stressed in Chapter 1, thinking that <truth> admits of a minimal analysis or that our mastery of the concept is relatively basic does not commit us to thinking that truth has no substantive nature; but, as deflationists have long emphasised, we do not need to postulate a substantive property ascribed by the predicate *to explain its expressive function*. If <truth> is governed by (unrestricted) (T-In) and (T-Out), then with some further plausible assumptions we can explain how it functions as a fully general device for endorsement.

Can this minimal analysis, however, be rendered consistent with the radical diversity in the metaphysics of truth that I have defended? One might think that ‘is true’ is context-sensitive. By comparison, there is no “absolute” sense of the word ‘tall’ such that something can be said to be tall, full-stop. Which things get to count as tall depends on context. Ajax is tall for a cat, but he is not tall for an animal. ‘True’, of course, is not a gradable adjective, but perhaps which property it ascribes also depends on context, and (T-In) and (T-Out) tell us about its *character*. This isn’t obviously a non-starter. For instance, perhaps my use of ‘true’ in ‘Whatever you just said was true’ picks out weak super-stability if you just said ‘Eating meat is wrong’, but  $T_{\&1}$  if you just said ‘Eating meat is normal but wrong’.<sup>2</sup> It does mean that ‘is true’ “equivocates” in principles like ‘A conjunction is true

<sup>1</sup> Lynch (2001: 726; 2004: 389; 2009: 57), Cotnoir (2013b: 563, fn.2).

<sup>2</sup> I’m here treating shifts in extension between contexts as being equivalent to ascribing different properties in different contexts.

iff its conjuncts are true’, but it’s not obvious that this immediately frustrates the expressive function of the truth predicate here (which, I suggested (§6.2.2), may be to highlight a *structural* constraint). However, it’s less straightforward to say what it ascribes when ‘true’ is used to endorse truthbearers apt for *different* properties, as it potentially does in ‘Everything you said yesterday is true’. Perhaps it ascribes a disjunctive property, depending on which truthbearers are being endorsed. Fortunately, however, we do not have to go context-sensitive here.

Suppose we think that ‘is true’ ascribes an unrestricted truth property,  $T_U$ . There are a few options here for rendering this consistent with the diversity of truth I have defended. First, one might endorse a kind of hybrid *intersectional* pluralism (§1.3.2), where  $T_U$  is a deflationary property. That is, one might think that ‘is true’ ascribes a deflationary property for which every truthbearer is apt, but that each truthbearer is *also* apt for an inflationary truth property; and *which* inflationary truth property depends on its content and logical form, as I have argued. If  $p$  is apt for inflationary property  $T_i$ , then ( $T_i$ -In) and ( $T_i$ -Out) will be valid for  $p$ ; and since ( $T_U$ -In) and ( $T_U$ -Out) are likewise valid for  $p$ ,  $p$  cannot be  $T_i$  without being  $T_U$ , nor *vice versa*. Nonetheless, there is no *substantive* property in common among all the truths; which is relevant for all truthbearers. This would be my preferred route.<sup>3</sup>

Alternatively, one might think that ‘ $T_U$ ’ denotes a substantive property, which stands in a one/many relation to the form- and domain-restricted properties, as in Lynch’s functionalism or Edwards’s “simple determination” pluralism.<sup>4</sup> An in-between option here would be a kind of radical alethic disjunctivism, where the unrestricted truth property is defined via infinite disjunction.<sup>5</sup> Note that this would not, given the argument of Chapter 6, render the form-restricted properties redundant; merely postulating an unrestricted truth property does nothing to solve the Problems of “Mixed” Inferences and Compounds. I am pessimistic that a plausible story can be given about the one/many relationship, but I will not argue that here.

### 8.1.2 The Liar and Inconsistency

However, there is a Big Problem with thinking that there is an unrestricted truth property:

(Liar)            Liar is not true.

- |     |                             |                                |
|-----|-----------------------------|--------------------------------|
| (1) | Liar is true.               | <i>assumption for reductio</i> |
| (2) | ‘Liar is not true’ is true. | <i>from (1), substitution</i>  |
| (3) | Liar is not true.           | <i>from (2), (T-Out)</i>       |
| (4) | <i>Contradiction.</i>       | <i>from (1), (3)</i>           |

<sup>3</sup> I take it that this is a fleshing out of the view suggested most recently by Wright (2013).

<sup>4</sup> Especially Lynch (2009), Edwards (2011).

<sup>5</sup> For alethic disjunctivism, see especially Pedersen & Wright (2013b).

- |     |                             |                               |
|-----|-----------------------------|-------------------------------|
| (5) | Liar is not true.           | <i>reductio from (1)-(4)</i>  |
| (6) | 'Liar is not true' is true. | <i>from (5), (T-In)</i>       |
| (7) | Liar is true.               | <i>from (6), substitution</i> |
| (8) | <i>Contradiction.</i>       | <i>from (5), (7)</i>          |

On the assumption that Liar is either true or not true (which, it seems at first pass, it must be, unless negation is somehow restricted), we can conclude that it is both, which is a contradiction. Thinking that there is an unrestricted truth property is thus, to say the least, the source of a number of headaches. Indeed, a diagnosis of the liar paradox (and other associated paradoxes, like the Curry) that has been gaining traction recently is that <truth> is inconsistent, where for a concept to be inconsistent is for it to be “governed”, in a substantive sense, by inconsistent rules (or, equivalently, for these rules to be constitutive for the meaning of the relevant word).<sup>6</sup> In the case of <truth>, (T-In) and (T-Out) are taken to be inconsistent rules, as evidenced *inter alia* by the liar paradox.<sup>7</sup>

Why might this diagnosis be a comfort? After all, what’s *worrying* about the liar paradox is that we end up with a contradiction; isn’t postulating inconsistency in the concept merely conceding what’s so problematic? Compare the suggestion that truth is an inconsistent *property*. This, I take it, would be no comfort. The primary virtue of locating the inconsistency in the *concept* is precisely that it *avoids* the commitment to there being any inconsistency in the world as such; it is “merely” a semantic artefact, a consequence of our using a faulty representational device, and *not* representative of anything in the world.<sup>8</sup> So, whatever story we tell about the semantics of ‘is true’, it cannot ascribe a single truth property with unrestricted application.

But this is in at least *prima facie* tension with the claim that truth is a substantive property of explanatory and normative import. On the one hand, the liar paradox and the inconsistency proposal push us towards thinking there cannot be such a property, on pain of paradox; on the other, the Success Argument and Weak Super-Stability Limitation (I have argued) push us towards thinking that there must be.<sup>9</sup> There is, famously, far too little interaction between the literature on the metaphysics of truth and that concerned with the paradoxes. In what follows, I make significant

---

<sup>6</sup> E.g., Eklund (2002a, 2002b, 2008), Azzouni (2006, 2007), Patterson (2007a, 2007b, 2009), Scharp (2007a, 2007b, 2013). Note that some theorists, including Eklund, do not commit themselves to its being <truth> in particular that is the culprit, although others, including Scharp, definitely do.

<sup>7</sup> I am not here presenting this as an argument *for* the inconsistency of <truth>, but merely explaining the suggestion. Given the overwhelming multiplicity of reactions to the liar paradox, making the case for the inconsistency proposal is well beyond the scope of this chapter. (See, e.g., Scharp (2013) for a book-length defence.)

<sup>8</sup> In some ways, I suppose, like “Penrose stairs” (the impossible, never-ending staircases made famous by Escher): they *look* fine, but they’re representationally faulty. See the epigraph of this chapter. We might think of restricted instances of (T-In) and (T-Out) as like the parts of the staircase – “acceptable” representations of different truth properties – but that the unrestricted rules, like the picture as a whole, are jointly inconsistent.

<sup>9</sup> Strictly, of course, if I’m right then they don’t push us to thinking that there’s *a* substantive property, but that there are substantive *properties*. It is precisely my argument here that thinking that there are substantive properties with restricted application, of the kind I have defended, allows us to avoid paradox consistently with inflationism about truth. Also note that this tension is not immediately dissolved if it is indeterminate which property ‘is true’ ascribes (as on Eklund’s (2002b) view).



steps towards resolving this tension, by showing how, if truth has the diverse nature that I have argued that it has – i.e., if there is an infinite multiplicity of *different* truth properties, each of which has *restricted* application – then we can understand the world, including the explanatorily and normatively important truth properties, as classically consistent, even while the concept is inconsistent. The thought is that ‘is true’ is able to play its fully general expressive role, not in virtue of ascribing an unrestricted truth property, because there is no such property, but rather precisely in virtue of voicing an inconsistent concept.<sup>10</sup>

## 8.2 Why Infinite Liars are Better than One

### 8.2.1 Tarski and Cotnoir on Infinite Liars

One approach to the liar paradox, exemplified by orthodox Tarskian hierarchical proposals, is to deny that there is a single truth predicate<sup>11</sup> that has unrestricted application; i.e., is such that (T-In) and (T-Out) are valid for *every* truthbearer. Instead, there is an infinite multiplicity of different truth predicates – ‘T<sub>1</sub>’, ‘T<sub>2</sub>’, etc. – such that, for each, the analogous introduction and elimination rules (T<sub>i</sub>-In) and (T<sub>i</sub>-Out) are valid for different sets of truthbearers. This means we can formulate an infinite multiplicity of different liar sentences, where a liar sentence is any sentence that says of itself that it does not possess a truth property. We will say that a particular liar sentence *concerns* the truth property (or properties) that it says it does not possess. Paradox is then avoided by maintaining that the relevant truth property for any liar sentence is not among the truth properties it concerns.<sup>12</sup>

For example, according to Tarski, truth is language-relative: each language L<sub>i</sub> has its own truth predicate ‘T<sub>i</sub>’. Languages are then barred from containing their own truth predicate. So, take the truth predicate ‘T<sub>1</sub>’ relevant for some language L<sub>1</sub>. The liar sentence Liar<sub>1</sub>

(Liar<sub>1</sub>)            Liar<sub>1</sub> is not T<sub>1</sub>

is not a sentence of L<sub>1</sub>, since it uses the predicate ‘T<sub>1</sub>’, which is not a part of L<sub>1</sub>. As such, (T<sub>1</sub>-In) and (T<sub>1</sub>-Out) are not valid for Liar<sub>1</sub>, so we cannot derive a contradiction. Instead, Liar<sub>1</sub> will be a part of a further language, L<sub>2</sub>. L<sub>2</sub> will have its own truth predicate, ‘T<sub>2</sub>’. We can, of course, formulate a new liar sentence that concerns T<sub>2</sub>, Liar<sub>2</sub>; but since ‘T<sub>2</sub>’ cannot be a part of L<sub>2</sub> itself, Liar<sub>2</sub>

<sup>10</sup> On this point, compare Burgess (2014: 585-6).

<sup>11</sup> Tarski talked in terms of predicates, as does Cotnoir in this setting. I’m going to talk pretty interchangeably about predicates and properties. Part of the reason is that, where other theorists are determined to remain neutral on the metaphysics, I am determined not to – this is a virtue, not a vice. For my purposes, a predicate is a truth predicate just in case it ascribes a truth property, where a truth property is still understood to be a property such that the analogous rules to the (T-In) and (T-Out) rules are valid for some set of truthbearers (§1.4.2).

<sup>12</sup> I talk about predicates and properties relevant for *sentences* here merely for ease of presentation.

will in turn not be a part of  $L_2$ , so (T<sub>2</sub>-In) and (T<sub>2</sub>-Out) are not valid for  $Liar_2$ . Instead,  $Liar_2$  is part of a further language,  $L_3$ . And so on and so forth up the hierarchy.

It is not my purpose here to argue against the orthodox Tarskian proposal, but a primary objection is often that such proposals look *ad hoc*.<sup>13</sup> What, other than paradox, might ban a language from containing its own truth predicate? Importantly, it's difficult to independently motivate the language-relativisation move, or to square it with the kinds of explanatory and normative roles that truth plays, which are not language-relative in any sense I can make sense of.

A similar proposal has recently been defended by Cotnoir (2013a) using domain-restricted pluralism about truth. Domain-restricted pluralists, as is by now familiar, hold that different truth properties are relevant for truthbearers in different domains. Cotnoir thus suggests that the domain-restricted pluralist hold that, if a particular truth property  $T_i$  is relevant for truthbearers in domain  $D_i$ , then any liar sentence that concerns  $T_i$  falls into some domain  $D_j$ ,  $j \neq i$ . Since we can disjunctively formulate liar sentences that concern more than one truth property, this requires (i) infinite domains, (ii) infinite truth properties, (iii) that each liar sentence falls into a different domain, and (iv) no infinite disjunctions. (iv) is necessary, since infinite disjunction would allow us to formulate a liar sentence that says of itself that it is not  $T_1$  or  $T_2$  or ..., for *every* domain-restricted truth property. Since there is *ex hypothesi* no further domain for this liar sentence to fall into, we fall back into paradox. Cotnoir is therefore concerned to provide independent (though still (Curry) paradox motivated) reason to disallow infinite disjunctions.

Even granting this, however, with only the resources of the domain-restricted pluralist to hand, (i)-(iii) look quite implausible. While pluralists have endorsed a multiplicity of different domains, there is no indication that this proliferation is *infinite*. On the most prominent account of domain individuation (§7.1.1), this would require infinite different kinds of individuals, properties, and relations; is it plausible that there are such? It is difficult to see how an independent argument might go.<sup>14</sup> It is not as though there is a dense or continuous parameter of variation, of the kind there is between, say, different lengths (of which there are infinite) between tables, numbers, and goodness. *Even granting* that there are infinite domains, (ii) does not follow: the domain-restricted pluralist is committed to thinking that which truth property is relevant for a truthbearer depends on its domain, not to thinking that *every* domain has its own truth property. What reason might there be for thinking that there are infinite domain-restricted truth properties? There is, again, no dense or continuous parameter of variation between correspondence, coherence, superwarrant, and weak super-stability.

---

<sup>13</sup> Tarski himself thought that the liar paradox shows that our language is inconsistent, so was trying to build a consistent replacement.

<sup>14</sup> Cook (2011) suggests that there may be infinite domains (in his terms, “discourses”). He says ‘[t]here seem to be no good reasons for thinking that there are only finitely many distinct discourses’ (2011: 627), but I do not think that is where the burden of proof lies.

And *even granting both of these*, (iii) still looks not only unmotivated but implausible. On the most prominent account of domain-individuation, which domain a truthbearer falls into depends on the *kind* of thing it is about, and all liar sentences are about the same kind of thing! (It’s obvious that my own metasemantically-driven account of domain individuation is of no help here.)

Perhaps most problematically, *even granting all of (i)-(iv)*, the central contention that *no* liar sentence falls into the domain for which the truth property it concerns is relevant for amounts to nothing more than a promissory note. Even if there are infinite domains, infinite truth properties, and each liar sentence falls into a different domain, one of them might concern its relevant truth property, at which point it’s all for naught. How can we tell that this won’t happen? Do we have to look at each liar sentence in turn?

## 8.2.2 The Infinite Liars of Form-Restricted Pluralism

### 8.2.2.1 A First Case

With form-restricted pluralism we’re in the enviable position of *already* having – independently motivated – the infinite machinery that this kind of proposal requires. (Logical form, if you like, provides a continuous parameter of variation along which we get both different sets of truthbearers and different truth properties.) To recall, for the primary truthbearers, we let ‘ $T_A$ ’ label whatever the truth of atomics consists in (for the domain-restricted pluralist, this can be a disjunction of the domain-restricted properties conjoined with their relevant domain); schematically, the relevant truth properties for complex truthbearers are given thus:

$$\begin{aligned} T_{\neg}(\neg p) &\leftrightarrow \neg T_A(p). \\ T_{\&}(p \& q) &\leftrightarrow (T_A(p) \& (T_A(q))). \\ \text{Etc.} \end{aligned}$$

So the “complexity” of the truth property – i.e., the logical form of the right-hand side of the definitional biconditional – mimics the logical form of the truthbearers for which that property is relevant. What we need is a principled reason to think that no liar sentence concerns its relevant truth property.

It will be illuminating to see how our machinery copes with the liar sentence that concerns the fundamental truth property,  $T_A$ :

$$(\text{Liar}_A) \quad \text{Not: } \text{Liar}_A \text{ is } T_A.^{15}$$

---

<sup>15</sup> I continue to use English words like ‘not’ rather than the formal notation ‘ $\neg$ ’ here to make the argument easier to follow; but I put the ‘not’ at the beginning in this way, since it (i) makes which truthbearer is negated obvious, and (ii) makes things easier when we have double negations later on.

As the negation of an atomic sentence, the relevant truth property for  $\text{Liar}_A$  is  $T_{\neg 1}$ , and whether or not it has this property depends on whether or not the sentence it negates,  $\text{Friend}_A$ , has its relevant truth property:

(Friend<sub>A</sub>)       $\text{Liar}_A$  is  $T_A$ .

(A “Friend” sentence is one that ascribes a truth property to a liar sentence. They will be important moving forward, since any complex depends for its truth property on its components, any liar sentence is a negation, and any liar sentence negates a friend sentence; so liar sentences depend for their truth properties on the friend sentences they negate.)

$\text{Friend}_A$  is atomic, so its relevant truth property is  $T_A$ . If  $\text{Liar}_A$  is not  $T_A$ , then what  $\text{Friend}_A$  says is wrong, so  $\text{Friend}_A$  is not  $T_A$ . In this scenario,  $\text{Liar}_A$  negates a sentence that is not  $T_A$ , and so possesses its relevant truth property  $T_{\neg 1}$  by definition. This is a good result, since in this scenario  $\text{Liar}_A$  gets things right: it says, of itself, that it is not  $T_A$ , and indeed it is not. Paradox is avoided because  $T_A$  is not its relevant truth property.

If  $\text{Liar}_A$  is  $T_A$ , somehow, then paradox is avoided nonetheless. In this scenario, what  $\text{Friend}_A$  says is right, so  $\text{Friend}_A$  is  $T_A$ . In which case  $\text{Liar}_A$  negates a sentence that is  $T_A$ , and so fails to possess its relevant truth property  $T_{\neg 1}$  by definition. This is a good result, since in this scenario  $\text{Liar}_A$  gets things wrong: it says, of itself, that it is not  $T_A$ , when in fact it is.

Either way we avoid paradox simply because  $\text{Liar}_A$  is of one order of complexity greater than the kind of truthbearer for which the truth property it concerns,  $T_A$ , is relevant.<sup>16</sup> This means that ( $T_A$ -In) and ( $T_A$ -Out) are not valid for  $\text{Liar}_A$ .

The goal is to show that this generalises to all other liar sentences.

### 8.2.2.2 A Problem Case

On this front, however, we immediately encounter a difficulty with the liar sentence that concerns the truth property relevant for first-order negations,  $T_{\neg 1}$ :

(Liar<sub>¬1</sub>)      Not:  $\text{Liar}_{\neg 1}$  is  $T_{\neg 1}$ .

$\text{Liar}_{\neg 1}$  negates  $\text{Friend}_{\neg 1}$ :

(Friend<sub>¬1</sub>)       $\text{Liar}_{\neg 1}$  is  $T_{\neg 1}$ .

---

<sup>16</sup> There is then a further question which scenario *in fact* obtains. For what it’s worth, my own view is that  $\text{Liar}_A$  is not  $T_A$ . But this is simply a further question, as far as our purposes go. What’s important is that we avoid paradox either way.

On the face of it,  $\text{Friend}_{\neg 1}$  is atomic and  $\text{Liar}_{\neg 1}$  is a first-order negation; so it looks like  $(\Gamma_A\text{-In})$  and  $(\Gamma_A\text{-Out})$  are valid for  $\text{Friend}_{\neg 1}$ ; and  $(\Gamma_{\neg 1}\text{-In})$  and  $(\Gamma_{\neg 1}\text{-Out})$  are valid for  $\text{Liar}_{\neg 1}$ . (9)-(19) shows one way we can derive a contradiction on these assumptions. In intuitive terms, the problem is that  $\text{Liar}_{\neg 1}$  seems to concern the very truth property that is relevant for truthbearers of its own kind. Suppose that  $\text{Liar}_{\neg 1}$  is  $\Gamma_{\neg 1}$ . Then what  $\text{Friend}_{\neg 1}$  says is right, so  $\text{Friend}_{\neg 1}$  is  $\Gamma_A$ . But then  $\text{Liar}_{\neg 1}$  negates a sentence that is  $\Gamma_A$ , and by definition is *not*  $\Gamma_{\neg 1}$ . This is a contradiction. On the other hand, if  $\text{Liar}_{\neg 1}$  is not  $\Gamma_{\neg 1}$ , then what  $\text{Friend}_{\neg 1}$  says is wrong, so  $\text{Friend}_{\neg 1}$  is not  $\Gamma_A$ . But then  $\text{Liar}_{\neg 1}$  negates a sentence that is not  $\Gamma_A$ , and by definition *is*  $\Gamma_{\neg 1}$ . This too is a contradiction. We appear to have landed back in paradox.

- |      |  |  |
|------|--|--|
| (9)  | $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ .                              | <i>assumption for reductio</i>                             |
| (10) | ' $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ ' is $\Gamma_A$ .            | <i>from (9), <math>(\Gamma_A\text{-In})</math></i>         |
| (11) | $\text{Friend}_{\neg 1}$ is $\Gamma_A$ .                                   | <i>from (10), substitution</i>                             |
| (12) | 'Not: $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ ' is $\Gamma_{\neg 1}$ . | <i>from (9), substitution</i>                              |
| (13) | Not: ' $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ ' is $\Gamma_A$ .       | <i>from (12), def. of '<math>\Gamma_{\neg 1}</math>'</i>   |
| (14) | Not: $\text{Friend}_{\neg 1}$ is $\Gamma_A$ .                              | <i>from (13), substitution</i>                             |
| (15) | <i>Contradiction.</i>  | <i>from (11), (14)</i>                                     |
| (16) | Not: $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ .                         | <i>reductio from (9)-(15)</i>                              |
| (17) | 'Not: $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ ' is $\Gamma_{\neg 1}$ . | <i>from (16), <math>(\Gamma_{\neg 1}\text{-In})</math></i> |
| (18) | $\text{Liar}_{\neg 1}$ is $\Gamma_{\neg 1}$ .                              | <i>from (17), substitution</i>                             |
| (19) | <i>Contradiction.</i>  | <i>from (16), (18)</i>                                     |

However, in the last chapter I argued that some *sentences* are non-primary; and, in particular, that some sentences are *non-canonical*, in the sense that their logical form does not match the logical form of the relevant primary truthbearer. For non-canonical sentences, we cannot read the relevant truth property off of their logical form. In particular, I suggested that some simple linguistic entities (words) plausibly voice complex concepts. If this is plausible for any words, it is plausible for those that explicitly have complex definitions, like 'badwork' and 'gbjag' from the last chapter. Significantly, the predicates that we have introduced to denote our complex truth properties are *precisely* of this kind.

This opens up the possibility that the liar sentence  $\text{Liar}_{\neg 1}$  is non-canonical, since ' $\Gamma_{\neg 1}$ ' arguably voices a complex concept. Suppose this is right, and  $\text{Liar}_{\neg 1}$  voices a complex proposition.<sup>17</sup> Deriving a contradiction here relied on the assumptions that  $(\Gamma_A\text{-In})$  and  $(\Gamma_A\text{-Out})$  are valid for

---

<sup>17</sup> As discussed in the last chapter, we need not think in terms of concepts and propositions here, but it is easier to talk in terms of a particular proposal; and this, in any case, is my preferred proposal.

Friend<sub>¬1</sub>, and (T<sub>¬1</sub>-In) and (T<sub>¬1</sub>-Out) are valid for Liar<sub>¬1</sub>. If Liar<sub>¬1</sub> (and Friend<sub>¬1</sub>) are non-canonical, this derivation will not go through.

What, then, are the relevant canonical “translations” of Liar<sub>¬1</sub> and Friend<sub>¬1</sub>, as it were? (A canonical “translation” will be a sentence with the same meaning that is canonical.) Friend<sub>¬1</sub> predicates ‘T<sub>¬1</sub>’ of Liar<sub>¬1</sub>, and ‘T<sub>¬1</sub>’ has the complex definition:

$$T_{\neg 1}(\neg p) \leftrightarrow \neg T_A(p).$$

Consider the following step-by-step translation:

(Friend<sub>¬1</sub>)      Liar<sub>¬1</sub> is T<sub>¬1</sub>.  
                          ‘Not: Liar<sub>¬1</sub> is T<sub>¬1</sub>’ is T<sub>¬1</sub>.  
                          Not: ‘Liar<sub>¬1</sub> is T<sub>¬1</sub>’ is T<sub>A</sub>.  
 (FRIEND<sub>¬1</sub>)    Not: Friend<sub>¬1</sub> is T<sub>A</sub>.

The second line is obtained from the first, and the last from the penultimate, by substitution. The third line is obtained from the second in accordance with the definition of ‘T<sub>¬1</sub>’. So at least by the ordinary lights of translation, it looks like what I have labelled ‘FRIEND<sub>¬1</sub>’ is a good translation of Friend<sub>¬1</sub>; that the two are synonymous.<sup>18</sup>

Similarly for Liar<sub>¬1</sub>: we might simply translate this by negating FRIEND<sub>¬1</sub>, but we can provide an analogous step-by-step translation as follows:

(Liar<sub>¬1</sub>)            Not: Liar<sub>¬1</sub> is T<sub>¬1</sub>.  
                          Not: ‘Not: Liar<sub>¬1</sub> is T<sub>¬1</sub>’ is T<sub>¬1</sub>.  
                          Not: Not: ‘Liar<sub>¬1</sub> is T<sub>¬1</sub>’ is T<sub>A</sub>.  
 (LIAR<sub>¬1</sub>)         Not: Not: Friend<sub>¬1</sub> is T<sub>A</sub>.

Again, the second and final lines are obtained by substitution, and the third from the definition of ‘T<sub>¬1</sub>’.

For all this, FRIEND<sub>¬1</sub> and LIAR<sub>¬1</sub> may not be canonical sentences. In particular, if ‘T<sub>A</sub>’ voices a disjunction of domain-restricted truth properties, then we have as much reason to doubt the canonicity of FRIEND<sub>¬1</sub> and LIAR<sub>¬1</sub> as we do their original forms. However, this turns out not to matter for our purposes, for reasons I’ll explain below. So, for the time being let’s imagine that they are canonical.

FRIEND<sub>¬1</sub> negates a sentence we can call S:

---

<sup>18</sup> One might worry that Friend<sub>¬1</sub> talks about Liar<sub>¬1</sub>, while FRIEND<sub>¬1</sub> talks about Friend<sub>¬1</sub> – how can they be good translations if they’re talking about different things? The answer is that Friend<sub>¬1</sub> is *part* of Liar<sub>¬1</sub>.

(S) Friend<sub>¬1</sub> is T<sub>A</sub>.

The relevant truth property for S, qua atomic sentence, is T<sub>A</sub>.<sup>19</sup> As a first-order negation, the relevant truth property for FRIEND<sub>¬1</sub> is not T<sub>A</sub>, but T<sub>¬1</sub>; and the relevant truth property for LIAR<sub>¬1</sub> as a second-order negation is not T<sub>¬1</sub> but T<sub>¬2</sub>:

$T_{¬2}('¬p') \leftrightarrow ¬T_{¬1}('p')$ ; or, equivalently:

$T_{¬2}('¬p') \leftrightarrow ¬¬T_A('p')$ .

(Note that I am only talking about the relevant truth properties for the “translations” of Friend<sub>¬1</sub> and Liar<sub>¬1</sub>, not for the sentences themselves. Their truth will be understood derivatively, in accordance with primalism (§1.1.2).)

We now have three crucial truthbearers in play, with three different relevant truth properties. Given the logical relations between them, which truth properties each possesses is dependent on the others:

		Relevant truth property:
(S)	Friend <sub>¬1</sub> is T <sub>A</sub> .	T <sub>A</sub> .
(FRIEND <sub>¬1</sub> )	Not: Friend <sub>¬1</sub> is T <sub>A</sub> .	T <sub>¬1</sub> .
(LIAR <sub>¬1</sub> )	Not: Not: Friend <sub>¬1</sub> is T <sub>A</sub> .	T <sub>¬2</sub> .

There are two possible scenarios.

*Scenario 1.* Suppose first that Friend<sub>¬1</sub> is T<sub>A</sub>. S gets things right, so S is T<sub>A</sub>. By definition, then, FRIEND<sub>¬1</sub> is not T<sub>¬1</sub> and LIAR<sub>¬1</sub> is T<sub>¬2</sub>.

FRIEND<sub>¬1</sub> does not have its relevant truth property. This is a good result, since FRIEND<sub>¬1</sub> says that Friend<sub>¬1</sub> is not T<sub>A</sub>, which we are supposing it is. LIAR<sub>¬1</sub>, on the other hand, does have its relevant truth property. This too is a good result, since if FRIEND<sub>¬1</sub> gets things wrong then LIAR<sub>¬1</sub>, as its negation, gets things right.

*Scenario 2.* Suppose instead that Friend<sub>¬1</sub> is not T<sub>A</sub>. S gets things wrong, so S is not T<sub>A</sub>. By definition, then, FRIEND<sub>¬1</sub> is T<sub>¬1</sub> and LIAR<sub>¬1</sub> is not T<sub>¬2</sub>.

FRIEND<sub>¬1</sub> has its relevant truth property. This is a good result, since FRIEND<sub>¬1</sub> says that Friend<sub>¬1</sub> is not T<sub>A</sub>, which we are supposing it is not. LIAR<sub>¬1</sub>, on the other hand, does not have its relevant truth property. This too is a good result, since if FRIEND<sub>¬1</sub> gets things right then LIAR<sub>¬1</sub>, as its negation, gets things wrong.

---

<sup>19</sup> I will talk for simplicity in terms of the relevant truth properties for the sentences as the sentences are, by hypothesis, of the same logical form as the primary truthbearers, but really this should be understood as meaning the relevant truth properties for the relevant primary truthbearer, whatever it is.

On either assumption, then, there is absolutely no contradiction when it comes to FRIEND<sub>-1</sub> and LIAR<sub>-1</sub>.

### 8.2.2.3 Generalising the Solution

If this is right, then we avoid paradox for Liar<sub>-1</sub> in much the same way as we did for Liar<sub>A</sub>: its canonical translation is one order of complexity greater than the truthbearers for which the truth property it concerns is relevant.

Furthermore, from the structure of the response we can see that it will generalise to any other liar sentence, and in turn will apply even if S, FRIEND<sub>-1</sub>, and LIAR<sub>-1</sub> are non-canonical themselves.

First, the canonical translation (or “more canonical” translation) of the predication of any complex truth property (of a Friend sentence) will be given by the right-hand side of the definitional biconditional, which only mentions T<sub>A</sub> (with the relevant change in named truthbearer). As mentioned (§6.2.1), the right-hand side of such definitional biconditionals will mimic the logical form of the truthbearers for which the truth property in question is relevant. This is because the relevant truth property for any complex is grounded in the truth properties of its components in just the way its logical form reveals. ‘p ∨ q’ is true just in case either ‘p’ is true or ‘q’ is true; so the truth property relevant for (first-order) ‘p ∨ q’ is the one it has just in case either ‘p’ is T<sub>A</sub> or ‘q’ is T<sub>A</sub>. So, the right-hand side of the definitional biconditionals, which give us the canonical translations of the Friend sentences, are of the complexity for which the truth property itself is relevant.

In turn, any liar sentence will be the negation of a truth ascription (to itself). So the canonical translation of any liar sentence will be *the negation* of a complex for which the truth property the liar sentence concerns is relevant; which entails that this translation of the liar sentence *itself* will be *one order of complexity greater* than the truthbearers for which the property it concerns is relevant. Thus the truth property relevant for any particular liar sentence will be distinct from the one(s) the liar sentence itself concerns.

For this reason, it does not matter if FRIEND<sub>-1</sub> and LIAR<sub>-1</sub> are non-canonical. What matters for the proposal is that the canonical translation of the liar sentences are *more* complex than the truthbearers for which the properties they concern are relevant. Should it turn out that, say, the proposition expressed by Liar<sub>-1</sub> is *even more* complex than revealed by LIAR<sub>-1</sub>, the proposal still runs.

This is no small result. For Tarski, this stage of the proposal was stipulated; for Cotnoir, it was a promissory note. We’ve shown that it follows, given form-restricted pluralism, on the assumption that the liar sentences formulated using predicates that ascribe complex truth properties mask the underlying complexity of the relevant primary truthbearer. This is a substantive



assumption, but it's not implausible, and it's not absent independent motivation; certainly within the general framework developed here.<sup>20</sup>

### 8.2.3 Revenge?

The threat of revenge for responses to the liar that rely on postulating a multiplicity of truth predicates comes primarily from disjunctions truth predicates. If 'Not: Liar is  $T_1$ ' is not apt for  $T_1$  but  $T_2$ , then what about 'Not: (Liar<sub>1V2</sub> is  $T_1$  or Liar<sub>1V2</sub> is  $T_2$ )'? And so on. This is why proposals of this style *require* infinite truth predicates, though of course I have argued that we ought to endorse infinite truth properties anyway. But what about infinite disjunction? Must we, with Cotnoir, rule it out?

*Mere* disjunction of truth predicates does not pose much of a problem here, for a reason we've already had cause to mention. As long as the truthbearer the (canonical translation of) the liar sentence negates is *at least* as complex as those truthbearers for which the truth property it concerns is relevant, we avoid paradox. Disjoining truth predicates merely adds to the complexity. If the liar sentences that negate ascriptions of single truth properties are guaranteed to be more complex than the truthbearers for which that truth property is relevant, then those that negate *disjunctions* of such properties are guaranteed to be so too. This is structurally analogous to Tarski's prohibition on a language containing its own truth predicate: where introducing a new truth predicate introduced a new language for Tarski, for the form-restricted pluralist, the increased complexity comes along in the structure of the truth properties.

I am no fan of infinite disjunction, and believe its rejection can be independently motivated. Nonetheless, there is room to allow for infinite disjunction within the present framework. The trick is to realise that infinitely complex truthbearers, if they exist, can nonetheless have different logical form: there are other infinite operations, e.g., infinite conjunctions; but, more importantly, an infinite disjunction might disjoin infinite atomic sentences, or infinite first-order negations, or any possible combination of truthbearers of different complexity. If the form-restricted pluralist allows infinite disjunction, then, there is no commitment to thinking that all infinite disjunctions are apt for the same truth property, simply because they are all infinite and all disjunctions. Rather, the relevant truth property for an infinite disjunction will depend on the logical form of its disjuncts, in just the same way that the relevant truth property for *any* logical complex depends on the logical form of its components (§6.2.1).

With that in mind, consider the liar sentence formulated using a truth predicate, ' $T_{\text{finite}}$ ', defined by disjoining *all* the truth predicates relevant for finite truthbearers, i.e., ' $x$  is  $T_{\text{finite}}$  iff  $x$  is  $T_A$

---

<sup>20</sup> It is worth mentioning here that form-restricted pluralism has equally fruitful application to the Curry paradox. I leave development to another occasion, but it should be apparent from the discussion of the liar how it will go. The problem case is  $T_{\rightarrow 1}$ , rather than  $T_{\neg 1}$ ; and we avoid paradox if uses of ' $T_{\rightarrow 1}$ ' are non-canonical.

or  $T_{\neg 1}$  or  $T_{\wedge 1}$  or  $T_{\neg 2}$  or...'. The canonical form of this liar sentence will be the negation of an infinite disjunction,<sup>21</sup> so none of those truth properties disjoined will be the relevant for it (given that they are only relevant for finite truthbearers); there is thus no threat of paradox here.

Of course, we can take the truth property that *is* relevant for this liar sentence – call it ' $T_{\omega 1}$ ' – and “add” it to the infinite disjunction. The canonical form of this truth property will itself be infinitely complex; this follows from our observation that the canonical form of an ascription of a complex truth property is at least as complex as those truthbearers for which it is relevant, which flows from the fact that the complex truth properties are grounded in  $T_A$ . If we thus “add”  $T_{\omega 1}$  to the infinite disjunction  $T_{\text{finite}}$ , then we will have an infinite disjunction with an infinitely complex disjunct, which is thus of different logical form to  $T_{\text{finite}}$  itself. The liar sentence formulated using the truth predicate so defined will thus be of *different* logical form to the one that concerns  $T_{\text{finite}}$ ; so  $T_{\omega 1}$  will not be its relevant truth property. And this dialectic iterates. There is thus always principled scope for the form-restricted pluralist to deny that a liar sentence will concern the truth property relevant for itself, even if we allow infinite disjunction. A similar story will run with regards to quantification over the truth properties: any such quantification over any particular set of truth properties will itself be apt for some *other* truth property.

My own preferred route, as suggested, is to reject infinite disjunction; but it is a virtue of the account that it is in principle consistent with such logical operations.

Of course, in a certain sense, we *can* arrive back at paradox. Consider the predicate, ‘has its relevant truth property’, or some such. But my proposal is *not* that we can’t define an unrestricted truth predicate; on the contrary, the English predicate ‘is true’ *is* just a predicate. On the inconsistency proposal, however, this is precisely *because* it voices an inconsistent concept. The picture I have been arguing for is one in which it is the assumption that there is an unrestricted truth property – a property for which every truthbearer is apt – that gives rise to paradox. On my view, there is no such property; but there is still a substantive story to tell about what the truth of any particular truthbearer consists in. This allows us to understand truth as an explanatorily and normative important part of our world, while explaining the liar paradox as resulting from ‘is true’ voicing an inconsistent concept.

## Conclusion

At the atomic level, truth consists in substantive properties: correspondence and weak super-stability. For complex truthbearers, truth consists in complex properties grounded in these properties. I have shown how, on this view, we can avoid the liar paradox, by developing a proposal that is structurally analogous to the orthodox Tarskian hierarchical proposal, but which uses

---

<sup>21</sup> There are difficulties in understanding how we can perform logical operations on infinite complexes. But that is all the more reason to be sceptical of infinite disjunction.

machinery independently motivated in the previous chapters. In this way, we can understand truth as a substantive part of a classically consistent world.

The Problem of Mixed Generalisations – or, generally, the expressive function of the truth predicate – gives us good reason to endorse an unrestricted *concept* of truth. But the liar paradox gives us good reason to think that there is no unrestricted truth property determinately ascribed by such a predicate. Indeed, I am tempted to say that it is only *because* ‘is true’ voices an inconsistent concept that it can play its role as a *fully general* device for endorsement. Consider, ‘Whatever you just said is true’. This endorses *whatever* you just said, including if what you just said is ‘Whatever you’re about to say is not true’.

This still leaves the job of saying how the inconsistent, unrestricted concept <truth> relates to the substantive properties. A detailed story here would require developing a semantics for inconsistent predicates, which is not a task I can undertake here. However, it is part of the picture that the truth predicate, despite voicing an inconsistent concept, *tracks* the substantive properties of explanatory and normative import that I have defended throughout the thesis. When talking about atomic moral truthbearers, for instance, the truth predicate tracks weak super-stability; when talking about truthbearers within the scope of the Success Argument, it tracks correspondence.<sup>22</sup> This can feature as part of an explanation as to why truth has always seemed to be a property of such import despite the predicate’s voicing an inconsistent concept.

By analogy, consider pre-Einsteinian use of the word ‘mass’. It turns out that some of the features Newtonian mechanics attributed to mass are true of *relativistic mass*, while others are true of *rest mass*. Our use of the term was thus of explanatory import, even if there was no one property that ‘mass’ determinately ascribed, because it tracked these different properties in different contexts (and in other contexts they coincided).<sup>23</sup> Of course, if I am right then there are (infinitely) many more properties in play here. But the analogy should be illuminating as to how our use of a predicate without unrestricted referent might nonetheless have been useful in virtue of its use tracking real properties in certain contexts.

---

<sup>22</sup> Think again, here, of the analogy with the Penrose stairs (fn.8).

<sup>23</sup> The canonical discussion of the ‘mass’ case is Field (1973); Scharp (2013: 37-8) also talks about it in the context of an inconsistency theory of the truth concept. For a similar analogy, consider Kitcher (1978) on ‘dephlogisticated air’, which in some contexts tracked the presence of oxygen, and hence was explanatorily useful, despite being a theoretical term in a false theory.

---

# CONCLUDING REMARKS

---

‘...which now as I catch up with myself, where I’ve gone and where I haven’t gone and what I better get back to, may very well have not been a pun at all but plain and simple just the bifurcation of truth, with an ampersand tossed in for unity.’  
Danielewski (2000: 145, fn.189).

In the preceding chapters, I have argued for the following theses:

- (i) *Anti-Ontological Drivers.* Extant ontologically-driven versions of the scope problem offer only weak motivation for domain-restricted anti-/realist truth pluralism. (Ch.2.)
- (ii) *Anti-Deflationism.* The deflationary construal of the principle that true belief facilitates practical success collapses the distinction between coincidental and non-coincidental instances of success; therefore, the deflationist cannot explain non-coincidental instances of success. (Ch.3.)
- (iii) *Pro-Correspondence.* Only a realist truth property can explain non-coincidental instances of success. (§4.1.1.)
- (iv) *Limited Scope.* Moral truths, expressivistically-conceived, fall outside the scope of the Success Argument. (§§4.1.2-4.2.3)
- (v) *Epistemic Constraint.* Metaethical expressivists are committed to the moral truths being epistemically constrained. (§§5.1-5.2.)
- (vi) *Anti-Smugness.* Egan’s “Smugness” worry requires that there is some admirable belief set that falls outside my network. This looks implausible, and it is incumbent on the would-be objector to show otherwise. (§5.3.)
- (vii) *Pro-Weak Super-Stability.* Moral truth consists in weak super-stability. (§§4.3-5.4)
- (viii) *Anti-Mixing Problems.* The Problems of “Mixed” Inferences and Compounds merely articulate structural constraints on the metaphysics of truth. (§6.1.)
- (ix) *Anti-Monism.* Endorsing monism about truth does not guarantee that one’s metaphysics satisfies the constraints of (viii). (§6.1.)
- (x) *Form-Restricted Pluralism.* The truth of a logically complex truthbearer consists in a property distinct from, but grounded in, the truth properties relevant for its components. (§6.2.)
- (xi) *Non-Primary Atomics.* Some atomic sentences are non-primary truthbearers, whose truth is to be understood in terms of the truth of a complex primary truthbearer. (Ch.7.)

- (xii) *Anti-Mixed Atomics*. Mixed atomic sentences in particular are non-primary. The relevant primary truthbearers are complexes composed out of unmixed atomics. (§7.2.)
- (xiii) *Unrestricted Concept*. The expressive function of the truth predicate is to be explained in terms of its voicing an unrestricted concept, which does not ascribe an unrestricted property if this concept is inconsistent. (§8.1.)
- (xiv) *Liar Dissolution*. Given (x) and (xi), even if <truth> is inconsistent, truth might nonetheless be an explanatorily and normatively important part of a classically consistent world. (§8.2.)

I have also suggested:

- (xv) *Frege-Geach*. An inflationary theory of moral truth may provide a new fast-track solution to the Frege-Geach Problem. (§4.2.)

*Anti-Ontological Drivers* is primarily of intra-pluralist import. Domain-restricted pluralists are better off offering direct *arguments for* thinking truth consists in one property here and another property there, and not relying on an intuitive “plausibility” or “implausibility” of one or another theory of truth in a particular domain. *Anti-Smugness* is likewise of intra-expressivist import, as is *Frege-Geach*.

The rest, taken together, constitute a complex pluralist metaphysics of truth. The truth predicate voices an unrestricted concept, which does not ascribe an unrestricted property because it is inconsistent (*Unrestricted Concept*). Nonetheless, the truth of some atomic truthbearers consists in correspondence, which can be used to explain the practical success of certain actions performed by rational agents (*Anti-Deflationism, Pro-Correspondence*); while the moral truths are epistemically constrained, and their truth consists in weak super-stability (*Limited Scope, Epistemic Constraint, Pro-Weak Super-Stability*). Complex truthbearers are apt for distinct truth properties that are grounded in the correspondence and weak super-stability (or otherwise) of their components (*Anti-Mixing Problems, Anti-Monism, Form-Restricted Pluralism*). This allows us to understand how truth might nonetheless be an explanatorily and normatively important part of the world, even though the predicate voices an inconsistent concept (*Non-Primary Atomics, Liar Dissolution*).

I find the whole attractive (I would, wouldn't I), but very many of the elements are detachable. At the atomic level, one might buy *Anti-Deflationism* without buying *Pro-Correspondence*; or *Pro-Correspondence* without buying *Epistemic Constraint*; or *Epistemic Constraint* without buying *Pro-Weak Super-Stability*. Importantly, while *Liar Dissolution* does rely on *Form-Restricted Pluralism* and *Non-Primary Atomics*, one can buy these without buying pluralism at the atomic level (e.g., both *Pro-Correspondence* and *Pro-Weak Super-Stability*).

*Anti-Monism* is an important lesson for anyone interested in providing a comprehensive metaphysics of truth. Even if one does not buy my particular argument *for* domain-restricted pluralism, *Anti-Mixing Problems*, *Form-Restricted Pluralism*, *Anti-Mixed Atomics*, and *Unrestricted Concept* show how the pluralist ought to respond to the most prominent objections.

There are still many questions left to answer, such as: Which truths, precisely, fall within the scope of the Success Argument? What is the nature of truth for those non-expressivist truthbearers outside its scope? Is the scope of the argument of Chapter 5 just moral discourse, or does it extend to any expressivist discourse? (Would it extend to aesthetic discourse? Or would our account of fallibility be different there? Would the idea of stability within different frameworks offer a new way of combining meta-aesthetic expressivism with relativism?) Can the truth-conditional metasemantic program for complex sentences of Chapter 4 be carried out? How is the idea that <truth> is inconsistent best understood? What is its semantics? What, in particular, is its relation to the infinite truth properties I have defended?

Related to this last question, also pressing is the *metametase* question: How should we explain the meaning of metasemantic discourse; that is, talk of meaning, truth, reference, and so on? Gibbard (2012), for instance, has developed an *expressivist* metasemantics, which includes an alethic expressivism. Is this consistent with what I've argued here? Does alethic expressivism entail that truth-ascriptions are apt for weak super-stability? How might we understand the idea that <truth> is inconsistent within the expressivist framework? Might the idea that the liar sentence expresses a “flip-flop” proposition (§5.2.5) be of use?

This thesis has concerned pluralism about truth. There are two key lessons that I want to finish by highlighting. First, I have developed a novel (and, given the argument of Chapter 2, I think stronger) motivation for domain-restricted anti-/realist pluralism. In providing the “direct” argument of Part II, I hope to exemplified a more compelling *style* of argument for pluralism, which I think future motivations should exemplify also.

Just as important, I think, has been the argument of Part III. My goal there was to show that pluralism does not give rise to new and embarrassing problems – far from it. The right metaphysics for complex truthbearers, I have suggested, dissolves these problems (and the liar paradox to boot), and is in an important sense *neutral* with regards to the nature of truth at the atomic level. If the arguments there are correct, I think this licences a kind of *methodological* pluralism, or at least a methodological *tolerance* for pluralism at the atomic level. This essentially frees the theorist from a monistic constraint, which can only make life easier.

## References

- Armstrong, D. 2004. *Truth and truthmakers*. Cambridge: Cambridge University Press.
- Asay, J. 2013. *The primitivist theory of truth*. Cambridge: Cambridge University Press.
- Asay, J. 2014. Against truth. *Erkenntnis*. **79**(1), pp.147-164.
- Asay, J. 2016. Putting pluralism in its place. *Philosophy and Phenomenological Research*. DOI: 10.1111/phpr.12303
- Ayer, A. 1936. *Language, truth and logic*. London: Penguin Books.
- Azzouni, J. 2006. *Tracking reason: proof, consequence, and truth*. Oxford: Oxford University Press.
- Azzouni, J. 2007. The inconsistency of natural languages: how we live with it. *Inquiry: An Interdisciplinary Journal of Philosophy*. **50**(6), pp.590-605.
- Baker, D. and Woods, J. 2015. How expressivists can and should explain inconsistency. *Ethics*. **125**(2), pp.391-424.
- Ball, B. 2017. Alethic pluralism and the role of reference in the metaphysics of truth. *Southern Journal of Philosophy*. **55**(1), pp.116-135.
- Barnard, R. and Horgan, T. 2006. Truth as mediated correspondence. *The Monist*. **89**(1), pp.28-49.
- Barnard, R. and Horgan, T. 2013. The synthetic unity of truth. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.180-196.
- Beall, J. 2000. On mixed inferences and pluralism about truth predicates. *Philosophical Quarterly*. **50**(200), pp.380-382.
- Beall, J. 2013. Deflated truth pluralism. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.323-338.
- Blackburn, S. 1984. *Spreading the word: groundings in the philosophy of language*. Oxford: Oxford University Press.
- Blackburn, S. 1992. Morality and thick concepts: through thick and thin. *Proceedings of the Aristotelian Society*. **66**(1), pp.285-299.
- Blackburn, S. 1993. *Essays in quasi-realism*. Oxford: Oxford University Press.
- Blackburn, S. 1996. Securing the nots: moral epistemology for the quasi-realist. In: Sinnott-Armstrong, W. and Timmons, M. eds. *Moral knowledge? New readings in moral epistemology*. Oxford: Oxford University Press, pp.82-100.
- Blackburn, S. 1998a. *Ruling passions: a theory of practical reasoning*. Oxford: Oxford University Press.
- Blackburn, S. 1998b. Wittgenstein, Wright, Rorty and minimalism. *Mind*. **107**(425), pp.157-181.
- Blackburn, S. 1999. Is objective moral justification possible on a quasi-realist foundation? *Inquiry: An Interdisciplinary Journal for Philosophy*. **42**(2), pp.213-227.
- Blackburn, S. 2001. *Think: a compelling introduction to philosophy*. Oxford: Oxford University Press.
- Blackburn, S. 2009. Truth and *a priori* possibility: Egan's charge against quasi-realism. *Australasian Journal of Philosophy*. **87**(2), pp.201-213.

- Blackburn, S. 2010. Truth, beauty and goodness. In: Schafer-Landau, R. ed. *Oxford Studies in Metaethics*. **5**(1), pp.295-314.
- Blackburn, S. 2013a. Deflationism, pluralism, expressivism, pragmatism. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.263-277.
- Blackburn, S. 2013b. Disentangling disentangling. In: Kirchin, S. ed. *Thick concepts*. Oxford: Oxford University Press, pp.121-135.
- Blackburn, S. 2015. Apologia pro vita sua. In: Johnson, R. and Smith, M. eds. *Passions and projections: themes from the philosophy of Simon Blackburn*. Oxford: Oxford University Press, pp.xv-xix.
- Boghossian, P. 2003. The normativity of content. *Philosophical Issues*. **13**(1), pp.31-45.
- Brandom, R. 1994. *Making it explicit: reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Brun, G. 2016. Explication as a method of conceptual re-engineering. *Erkenntnis*. **81**(6), pp.1211-1241.
- Burge, T. 1979. Individualism and the mental. *Midwest Studies in Philosophy*. **4**(1), pp.73-122.
- Burgess, A. 2014. Keeping 'true': a case study in conceptual ethics. *Inquiry: An Interdisciplinary Journal of Philosophy*. **57**(5-6), pp.580-606.
- Burgess, A. and Plunkett, D. 2013a. Conceptual ethics I. *Philosophy Compass*. **8**(12), pp.1091-1101.
- Burgess, A. and Plunkett, D. 2013b. Conceptual ethics II. *Philosophy Compass*. **8**(12), pp.1102-1110.
- Bykvist, K. and Hattiangadi, A. 2007. Does thought imply ought? *Analysis*. **67**(296), pp.277-285.
- Bykvist, K. and Hattiangadi, A. 2013. Belief, truth, and blindspots. In: Chan, T. ed. *The aim of belief*. Oxford: Oxford University Press, pp.80-99.
- Cappelen, H. [Forthcoming]. *Fixing language: conceptual engineering and the limits of revision*.
- Capps, D., Lynch, M., and Massey, D. 2008. A coherent moral relativism. *Synthese*. **166**(2), pp.413-430.
- Caputo, S. 2012. Three dilemmas for alethic functionalism. *Philosophical Quarterly*. **62**(249), pp.853-861.
- Carnap, R. 1950. *Logical foundations of probability*. Chicago: University of Chicago Press.
- Chalmers, D. 2011. Verbal disputes. *Philosophical Review*. **120**(4), pp.515-566.
- Cook, R. 2011. Alethic pluralism, generic truth, and mixed conjunctions. *Philosophical Quarterly*. **61**(244), pp.624-629.
- Cotnoir, A. 2009. Generic truth and mixed conjunctions: some alternatives. *Analysis*. **69**(3), pp.473-479.
- Cotnoir, A. 2013a. Pluralism and paradox. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University press, pp.339-350.
- Cotnoir, A. 2013b. Validity for strong pluralists. *Philosophy and Phenomenological Research*. **86**(3), pp.563-579.



- Cotnoir, A. and Edwards, D. 2015. From truth pluralism to ontological pluralism and back. *Journal of Philosophy*. **112**(3), pp.113-140.
- Damnjanovic, N. 2005. Deflationism and the success argument. *Philosophical Quarterly*. **55**(218), pp.53-67.
- Danielewski, M. 2000. *House of leaves*. 2nd ed. New York, NY: Random House.
- David, M. 1994. *Correspondence and disquotatation: an essay on the nature of truth*. Oxford: Oxford University Press.
- David, M. 2013. Lynch's functionalist theory of truth. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.42-68.
- David, M. 2016. The correspondence theory of truth. In: Zalta, E. ed. *The Stanford encyclopedia of philosophy (fall 2016 edition)*. URL = <<https://plato.stanford.edu/archives/fall2016/entries/truth-correspondence/>>.
- Davidson, D. 1990. The structure and content of truth. *Journal of Philosophy*. **87**(6), pp.279-328.
- Davidson, D. 1996. The folly of trying to define truth. *Journal of Philosophy*. **93**(6), pp.263-278.
- Devitt, M. 1981. *Designation*. New York, NY: Columbia University Press.
- Devitt, M. 1991. *Realism and truth*. 2nd ed. Princeton, NJ: Princeton University Press.
- Divers, J. and Miller, A. Why expressivists about value should not love minimalism about truth. *Analysis*. **54**(1), pp.12-19.
- Divers, J. and Miller, A. Platitudes and attitudes: a minimalist conception of belief. *Analysis*. **55**(1), pp.37-44.
- Dodd, J. 1999. There is no norm of truth: a minimalist reply to Wright. *Analysis*. **59**(4), pp.291-299.
- Dodd, J. 2000. *An identity theory of truth*. Basingstoke: Palgrave Macmillan.
- Dodd, J. 2013. Deflationism trumps pluralism! In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.298-322.
- Dreier, J. 1996. Expressivist embeddings and minimalist truth. *Philosophical Studies*. **83**(1), pp.29-51.
- Dreier, J. 2004. Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*. **18**(1), pp.23-44.
- Dretske, F. 1981. *Knowledge and the flow of information*. Cambridge, MA: The MIT Press.
- Dretske, F. 1988. *Explaining behavior*. Cambridge, MA: The MIT Press.
- Dretske, F. 1995. *Naturalizing the mind*. Cambridge, MA: The MIT Press.
- Dretske, F. 2001. Norms, history and the mental. *Royal Institute of Philosophy Supplement*. **49**(1), pp.87-104.
- Dummett, M. 1959. Truth. *Proceedings of the Aristotelian Society*. **59**(1), pp.141-162.
- Dummett, M. 1978. *Truth and other enigmas*. London: Duckworth.
- Dummett, M. 1993. *The seas of language*. Oxford: Oxford University Press.

- Dutilh Novaes, C. and Reck, E. 2017. Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese*. **194**(1), pp.195-215.
- Edwards, D. 2008. How to solve the problem of mixed conjunctions. *Analysis*. **68**(2), pp.143-149.
- Edwards, D. 2009. Truth-conditions and the nature of truth: re-solving mixed conjunctions. *Analysis*. **69**(4), pp.684-688.
- Edwards, D. 2011. Simplifying alethic pluralism. *Southern Journal of Philosophy*. **49**(1), pp.28-48.
- Edwards, D. 2012a. Alethic vs deflationary functionalism. *International Journal of Philosophical Studies*. **20**(1), pp.115-124.
- Edwards, D. 2012b. On alethic disjunctivism. *Dialectica*. **66**(1), pp.200-214.
- Edwards, D. 2013a. Naturalness, representation and the metaphysics of truth. *European Journal of Philosophy*. **21**(3), pp.384-401.
- Edwards, D. 2013b. Truth as a substantive property. *Australasian Journal of Philosophy*. **91**(2), pp.279-294.
- Edwards, D. 2013c. Truth, winning, and simple determination pluralism. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.113-122.
- Edwards, D. [Forthcoming]. *The metaphysics of truth*.
- Egan, A. 2007. Quasi-realism and fundamental moral error. *Australasian Journal of Philosophy*. **85**(2), pp.205-219.
- Egan, A. and Elga, A. 2005. I can't believe I'm stupid. *Philosophical Perspectives*. **19**(1), pp.77-93.
- Eklund, M. 2002a. Deep inconsistency. *Australasian Journal of Philosophy*. **80**(3), pp.321-331.
- Eklund, M. 2002b. Inconsistent languages. *Philosophy and Phenomenological Research*. **64**(2), pp.251-275.
- Eklund, M. 2008. Reply to Beall and Priest. *Australasian Journal of Logic*. **6**(1), pp.94-106.
- Eklund, M. 2010. Rejectionism about truth. In: Wright, C. and Pedersen, N. eds. *New waves in truth*. Basingstoke: Palgrave Macmillan, pp.30-44.
- Eklund, M. 2013. Evaluative language and evaluative reality. In: Kirchin, S. ed. *Thick concepts*. Oxford: Oxford University Press, pp.161-181.
- Eklund, M. 2014. Replacing truth? In: Burgess, A. and Sherman, B. eds. *Metasemantics: new essays on the foundations of meaning*. Oxford: Oxford University Press, pp.293-310.
- Eklund, M. 2015. Intuitions, conceptual engineering, and conceptual fixed points. In: Daly, C. ed. *The Palgrave handbook of philosophical methods*. London: Palgrave Macmillan, pp.363-385.
- Eklund, M. 2017. Inconsistency and replacement. *Inquiry: An Interdisciplinary Journal of Philosophy*. [Online]. [Accessed 5 March 2017]. Available from: <http://dx.doi.org/10.1080/0020174X.2017.1287918>
- Elstein, D. 2013. *Prescriptions and universalizability: a defence of Harean ethical theory*. Ph.D. thesis, University of Cambridge.

- Elstein, D. and Hurka, T. 2009. From thick to thin: two moral reduction plans. *Canadian Journal of Philosophy*. **39**(4), pp.515-535.
- Engel, P. 2013. In defence of normativism about the aim of belief. In: Chan, T. ed. *The aim of belief*. Oxford: Oxford University Press, pp.32-63.
- Evans, G. 1982. *The varieties of reference*. Oxford: Oxford University Press.
- Ferrari, F. 2016. Disagreement about taste and alethic suberogation. *Philosophical Quarterly*. **66**(264), pp.516-535.
- Field, H. 1972. Tarski's theory of truth. *Journal of Philosophy*. **69**(13), pp.347-375.
- Field, H. 1973. Theory change and the indeterminacy of reference. *Journal of Philosophy*. **70**(14), pp.462-481.
- Field, H. 1977. Logic, meaning and conceptual role. *Journal of Philosophy*. **74**(7), pp.379-409.
- Field, H. 1986. The deflationary conception of truth. In: MacDonald, G. and Wright, C. J. G. eds. *Fact, science and morality: essays on A. J. Ayer's Language, truth and logic*. Oxford: Basil Blackwell, pp.55-117.
- Field, H. 1994a. Deflationist views of meaning and content. *Mind*. **103**(411), pp.249-285.
- Field, H. 1994b. Disquotational truth and factually defective discourse. *Philosophical Review*. **103**(3), pp.405-452.
- Field, H. 2008. *Saving truth from paradox*. Oxford: Oxford University Press.
- Fodor, J. 1987. *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge, MA: The MIT Press.
- Fodor, J. 1990. *A theory of content and other essays*. Cambridge, MA: The MIT Press.
- Fodor, J. 1994. *The elm and the expert*. Cambridge, MA: The MIT Press.
- Gamester, W. [Forthcoming]. Truth: explanation, success, and coincidence. *Philosophical Studies*. [Online]. [Accessed 3 May 2017]. Available from: <https://doi.org/10.1007/s11098-017-0909-2>
- Gibbard, A. 1990. *Wise choices, apt feelings: a theory of normative judgement*. Oxford: Oxford University Press.
- Gibbard, A. 1992. Morality and thick concepts: thick concepts and warrant for feelings. *Proceedings of the Aristotelian Society*. **66**(1), pp.267-283.
- Gibbard, A. 2003. *Thinking how to live*. Cambridge, MA: Harvard University Press.
- Gibbard, A. 2005. Truth and correct belief. *Philosophical Issues*. **15**(1), pp.338-350.
- Gibbard, A. 2006. Normative properties. In: Horgan, T. and Timmons, M. eds. *Southern Journal of Philosophy*. **41**(s1), pp.141-157.
- Gibbard, A. 2012. *Meaning and normativity*. Oxford: Oxford University Press.
- Glüer, K. and Wikforss, Å. 2009. Against content normativity. *Mind*. **118**(469), pp.31-70.
- Glüer, K. and Wikforss, Å. 2013. Against belief normativity. In: Chan, T. ed. *The aim of belief*. Oxford: Oxford University Press, pp.100-122.

- Greenough, P. 2010. Deflationism and truth value gaps. In: Wright, C. D. and Pedersen, N. eds. *New waves in truth*. Basingstoke: Palgrave Macmillan, pp.115-125.
- Greenough, P. 2017. Conceptual Marxism and truth: *Inquiry* symposium on Kevin Scharp's *Replacing truth*. *Inquiry: An Interdisciplinary Journal of Philosophy*. [Online]. [Accessed 10 March 2017]. Available from: <http://dx.doi.org/10.1080/0020174X.2017.1287919>
- Grover, D. 1992. *A prosentential theory of truth*. Princeton, NJ: Princeton University Press.
- Grover, D., Kamp, J., and Belnap, N. 1975. A prosentential theory of truth. *Philosophical Studies*. **27**(1), pp.73-125.
- Halbach, V. and Horsten, L. 2006. Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic*. **71**(2), pp.677-712.
- Harcourt, E. and Thomas, A. 2013. Thick concepts, analysis, and reductionism. In: Kirchin, S. ed. *Thick concepts*. Oxford: Oxford University Press, pp.20-43.
- Haslanger, S. 1999. What knowledge is and what it ought to be: feminist values and normative epistemology. *Philosophical Perspectives*. **13**(1), pp.459-480.
- Haslanger, S. 2000. Gender and race: (what) are they? (What) do we want them to be? *Noûs*. **34**(1), pp.31-55.
- Haslanger, S. 2005. What are we talking about? The semantics and politics of social kinds. *Hypatia*. **20**(4), pp.10-26.
- Haslanger, S. 2006. Philosophical analysis and social kinds: what good are our intuitions? *Proceedings of the Aristotelian Society*. **80**(1), pp.89-118.
- Haslanger, S. 2012. *Resisting reality: social construction and social critique*. Oxford: Oxford University Press.
- Holton, R. 2000. Minimalism and truth-value gaps. *Philosophical Studies*. **97**(2), pp.137-168.
- Horgan, T. and Timmons, M. 2006. Cognitivist expressivism. In: Horgan, T. and Timmons, M. eds. *Metaethics after Moore*. Oxford: Oxford University Press, pp.255-298.
- Hornsby, J. 1997. Truth: the identity theory. *Proceedings of the Aristotelian Society*. **97**(1), pp.1-24.
- Horsten, L. 2009. Levity. *Mind*. **118**(471), pp.555-581.
- Horsten, L. 2011. *The Tarskian turn: deflationism and axiomatic truth*. Cambridge, MA: The MIT Press.
- Horton, M. and Poston, T. 2012. Functionalism about truth and the metaphysics of reduction. *Acta Analytica*. **27**(1), pp.13-27.
- Horwich, P. 1993. Gibbard's theory of norms. *Philosophy and Public Affairs*. **22**(1), pp.67-78.
- Horwich, P. 1994. The essence of expressivism. *Analysis*. **54**(1), pp.19-20.
- Horwich, P. 1998. *Truth*. 2nd ed. Oxford: Oxford University Press.
- Jackson, F. 1994. Realism, truth and truth aptness. *Philosophical Books*. **35**(3), pp.162-169.
- Jackson, F. 1998. *From metaphysics to ethics: a defence of conceptual analysis*. Oxford: Oxford University Press.
- Jackson, F., Oppy, G., and Smith, M. Minimalism and truth aptness. *Mind*. **103**(411), pp.287-302.

- Jackson, F. and Pettit, P. 1988. Functionalism and broad content. *Mind*. **97**(387), pp.381-400.
- Jackson, F. and Pettit, P. 1990. Program explanation: a general perspective. *Analysis*. **50**(2), pp.107-117.
- Jackson, F. and Pettit, P. 1992a. In defence of explanatory ecumenicalism. *Economics and Philosophy*. **8**(1), pp.1-21.
- Jackson, F. and Pettit, P. 1992b. Structural explanation in social theory. In: Lennon, K. and Charles, D. eds. *Reduction, explanation, and realism*. Oxford: Oxford University Press, pp.97-132.
- James, W. 1907. *Pragmatism: a new name for some old ways of thinking*. New York, NY: Longmans.
- Justus, J. 2012. Carnap on concept determination: methodology for philosophy of science. *European Journal for Philosophy of Science*. **2**(2), pp.161-179.
- King, J., Soames, S., and Speaks, J. 2014. *New thinking about propositions*. Oxford: Oxford University Press.
- Kirchin, S. 2010. The shapelessness hypothesis. *Philosophers' Imprint*. **10**(4), pp.1-28.
- Kirchin, S. ed. 2013. *Thick concepts*. Oxford: Oxford University Press.
- Kirkham, R. 1992. *Theories of truth: a critical introduction*. Cambridge, MA: The MIT Press.
- Kitcher, P. 1978. Theories, theorists and theoretical change. *Philosophical Review*. **87**(4), pp.519-547.
- Kitcher, P. 2002. On the explanatory role of correspondence truth. *Philosophy and Phenomenological Research*. **64**(2), pp.346-364.
- Köhler, S. 2014. Expressivism and mind-dependence. *Journal of Moral Philosophy*. **11**(6), pp.750-764.
- Köhler, S. 2015. What is the problem with fundamental moral error? *Australasian Journal of Philosophy*. **93**(1), pp.161-165.
- Kölbel, M. 2008. "True" as ambiguous. *Philosophy and Phenomenological Research*. **77**(2), pp.359-384.
- Kölbel, M. 2013. Should we be pluralists about truth? In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.278-297.
- Künne, W. 2003. *Conceptions of truth*. Oxford: Oxford University Press.
- Lando, T. 2017. Coincidence and common cause. *Noûs*. **51**(1), pp.132-151.
- Laurence, S. and Margolis, E. 1999. Concepts and cognitive science. In: Margolis, E. and Laurence, S. eds. *Concepts: core readings*. Cambridge, MA: MIT Press, pp.3-81.
- Leeds, S. 1978. Theories of reference and truth. *Erkenntnis*. **13**(1), pp.111-129.
- Leeds, S. 1995. Truth, correspondence, and success. *Philosophical Studies*. **79**(1), pp.1-36.
- Leeds, S. 2007. Correspondence truth and scientific realism. *Synthese*. **159**(1), pp.1-21.
- Lenman, J. 2014. Gibbardian humility: moral fallibility and moral smugness. *Journal of Value Inquiry*. **48**(2), pp.235-245.
- Lewis, D. 1975. Languages and language. In: Gunderson, K. ed. *Language, mind, and knowledge: Minnesota Studies in the Philosophy of Science*, **7**(1). Minneapolis, MN: University of Minnesota Press, pp.3-35.

- Lewis, D. 1983. New work for a theory of universals. *Australasian Journal of Philosophy*. **61**(4), pp.343-377.
- Lloyd, D. 1987. Mental representation from the bottom up. *Synthese*. **70**(1), pp.23-78.
- Lynch, M. 2001. A functionalist theory of truth. In: Lynch, M. ed. *The nature of truth: classic and contemporary perspectives*. Cambridge, MA: The MIT Press, pp.723-750.
- Lynch, M. 2004. Truth and multiple realizability. *Australasian Journal of Philosophy*. **82**(3), pp.384-408.
- Lynch, M. 2005. Alethic functionalism and our folk theory of truth: a reply to Cory Wright. *Synthese*. **145**(1), pp.29-43.
- Lynch, M. 2006. ReWrighting pluralism. *The Monist*. **89**(1), pp.63-84.
- Lynch, M. 2008. Alethic pluralism, logical consequence, and the universality of reason. *Midwest Studies in Philosophy*. **32**(1), pp.122-140.
- Lynch, M. 2009. *Truth as one and many*. Oxford: Oxford University Press.
- Lynch, M. 2012. The many faces of truth: a response to some critics. *International Journal of Philosophical Studies*. **20**(2), pp.255-269.
- Lynch, M. 2013a. Expressivism and plural truth. *Philosophical Studies*. **163**(2), pp.385-401.
- Lynch, M. 2013b. Three questions for truth pluralism. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.21-41.
- McDaniel, K. 2016. John M. E. McTaggart. In: Zalta, E. ed. *The Stanford encyclopedia of philosophy (winter 2016 edition)*. URL = <<https://plato.stanford.edu/archives/win2016/entries/mctaggart/>>.
- McDowell, J. 1981. Non-cognitivism and rule-following. In: Holtzman, S. and Leich, C. eds. *Wittgenstein: to follow a rule*. London: Routledge and Kegan Paul, pp.141-162.
- MacFarlane, J. 2014. *Assessment sensitivity: relative truth and its applications*. Oxford: Oxford University Press.
- McGee, V. 2005. Two conceptions of truth? – Comment. *Philosophical Studies*. **124**(1), pp.71-104.
- McGee, V. and McLaughlin, B. 1995. Distinctions without a difference. *Southern Journal of Philosophy*. **33**(S1), pp.203-251.
- Mackie, J. 1977. *Ethics: inventing right and wrong*. London: Penguin Books.
- Martin, C. 1996. How it is: entities, absences and voids. *Australasian Journal of Philosophy*. **74**(1), pp.57-65.
- Maher, P. 2007. Explication defended. *Studia Logica*. **86**(2), pp.331-341.
- Margolis, E. and Laurence, S. eds. 1999. *Concepts: core readings*. Cambridge, MA: The MIT Press.
- Millikan, R. 1984. *Language, thought and other biological categories*. Cambridge, MA: The MIT Press.
- Mumford, S. 2007. Negative truth and falsehood. *Proceedings of the Aristotelian Society*. **107**(1), pp.45-71.

- Neander, K. 2012. Teleological theories of mental content. In: Zalta, E. ed. *The Stanford encyclopedia of philosophy (spring 2012 edition)*. URL = <<https://plato.stanford.edu/archives/spr2012/entries/content-teleological/>>.
- Newhard, J. 2013. Four objections to alethic functionalism. *Journal of Philosophical Research*. **38**(1), pp.69-87.
- Newhard, J. 2014. Alethic functionalism, manifestation, and the nature of truth. *Acta Analytica*. **29**(3), pp.349-361.
- Newhard, J. 2017. Plain truth and the incoherence of alethic functionalism. *Synthese*. **194**(5), pp.1591-1611.
- Papineau, D. 1987. *Reality and representation*. Oxford: Basil Blackwell.
- Papineau, D. 1999. Normativity and judgement. *Proceedings of the Aristotelian Society*. **73**(1), pp.17-43.
- Papineau, D. 2013. There are no norms of belief. In: Chan, T. ed. *The aim of belief*. Oxford: Oxford University Press, pp.64-79.
- Patterson, D. 2007a. Inconsistency theories: the significance of semantic ascent. *Inquiry: An Interdisciplinary Journal of Philosophy*. **50**(6), pp.575-589.
- Patterson, D. 2007b. Understanding the liar. In: Beall, J. ed. *Revenge of the liar: new essays on the paradox*. Oxford: Oxford University Press, pp.197-224.
- Patterson, D. 2009. Inconsistency theories of semantic paradox. *Philosophy and Phenomenological Research*. **79**(2), pp.387-422.
- Patterson, D. 2010. Truth as conceptually primitive. In: Wright, C. D. and Pedersen, N. eds. *New waves in truth*. Basingstoke: Palgrave Macmillan, pp.13-29.
- Pedersen, N. 2006. What can the problem of mixed inferences teach us about alethic pluralism. *The Monist*. **89**(1), pp.103-117.
- Pedersen, N. 2010. Stabilizing alethic pluralism. *Philosophical Quarterly*. **60**(238), pp.92-108.
- Pedersen, N. 2012. True alethic functionalism? *International Journal of Philosophical Studies*. **20**(1), pp.125-133.
- Pedersen, N. 2014. Pluralism  $\times$  3: truth, logic, metaphysics. *Erkenntnis*. **79**(2), pp.259-277.
- Pedersen, N. and Edwards, D. 2011. Truth as one(s) and many: on Lynch's alethic functionalism. *Analytic Philosophy*. **52**(3), pp.213-230.
- Pedersen, N. and Wright, C. D. 2013a. Introduction. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.1-18.
- Pedersen, N. and Wright, C. D. 2013b. Pluralism about truth as alethic disjunctivism. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.87-112.
- Pedersen, N. and Wright, C. D. 2016. Pluralist theories of truth. In: Zalta, E. ed. *The Stanford encyclopedia of philosophy (spring 2016 edition)*. URL = <<https://plato.stanford.edu/archives/spr2016/entries/truth-pluralist/>>.

- Penrose, L. and Penrose, R. 1958. Impossible objects: a special type of visual illusion. *British Journal of Psychology*. **49**(1), pp.31-33.
- Perice, C. 1878. How to make our ideas clear. *Popular Science Monthly*. **12**(1), pp.286-302.
- Pettit, P. 1996. Realism and truth: a comment on Crispin Wright's *Truth and objectivity*. *Philosophy and Phenomenological Research*. **56**(4), pp.883-890.
- Plunkett, D. 2016. Conceptual history, conceptual ethics, and the aims of inquiry: a framework for thinking about the relevance of the history/genealogy of concepts to normative inquiry. *Ergo*. **3**(2), pp.27-64.
- Plunkett, D. and Sundell, T. 2013. Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*. **13**(23), pp.1-37.
- Putnam, H. 1975. The meaning of 'meaning'. In: Gunderson, K. ed. *Language, mind, and knowledge: Minnesota Studies in the Philosophy of Science*, **7**. Minneapolis, MN: University of Minnesota Press, pp.131-193.
- Putnam, H. 1978. *Meaning and the moral sciences*. London: Routledge and Kegan Paul.
- Quine, W. 1960. *Word and object*. Cambridge, MA: The MIT Press.
- Ramsey, F. 1927. Facts and propositions. *Proceedings of the Aristotelian Society*. **7**(1), pp.153-170.
- Ridge, M. 2014. *Impassioned belief*. Oxford: Oxford University Press.
- Ridge, M. 2015. I might be fundamentally mistaken. *Journal of Ethics and Social Philosophy*. **9**(3), pp.1-21.
- Roberts, D. 2013. It's evaluation, only thicker. In: Kirchin, S. ed. *Thick concepts*. Oxford: Oxford University Press, pp.78-96.
- van Roojen, M. 1996. Expressivism and irrationality. *Philosophical Review*. **105**(3), pp.311-335.
- Rupert, R. 1999. The best test theory of extension: first principle(s). *Mind and Language*. **14**(3), pp.321-355.
- Ryder, D. 2004. SINBaD neurosemantics: a theory of mental representation. *Mind and Language*. **19**(2), pp.211-240.
- Sainsbury, R. 1996. Crispin Wright: *Truth and objectivity*. *Philosophy and Phenomenological Research*. **56**(4), pp.899-904.
- Saul, J. 2006. Philosophical analysis and social kinds: gender and race. *Proceedings of the Aristotelian Society*. **80**(1), pp.119-143.
- Scharp, K. 2007a. Alethic vengeance. In: Beall, J. ed. *Revenge of the liar: new essays on the paradox*. Oxford: Oxford University Press, pp.272-319.
- Scharp, K. 2007b. Replacing truth. *Inquiry: An Interdisciplinary Journal of Philosophy*. **50**(6), pp.606-621.
- Scharp, K. 2013. *Replacing truth*. Oxford: Oxford University Press.
- Schroeder, M. 2008. *Being for: evaluating the semantic program of expressivism*. Oxford: Oxford University Press.



- Schroeder, M. 2010a. How to be an expressivist about truth. In: Wright, C. D. and Pedersen, N. eds. *New waves in truth*. Basingstoke: Palgrave Macmillan, pp.282-298.
- Schroeder, M. 2010b. *Non-cognitivism in ethics*. London: Routledge.
- Schroeder, M. 2014. Does expressivism have subjectivist consequences? *Philosophical Perspectives*. **28**(1), pp.278-290.
- Shepherd, J. and Justus, J. 2015. X-phi and Carnapian explication. *Erkenntnis*. **80**(2), pp.381-402.
- Shah, N. 2003. How truth governs belief. *Philosophical Review*. **112**(4), pp.447-482.
- Shah, N. and Velleman, D. 2005. Doxastic deliberation. *Philosophical Review*. **114**(4), pp.497-534.
- Shapiro, S. 2011. Truth, function and paradox. *Analysis*. **71**(1), pp.38-44.
- Sher, G. 1998. On the possibility of a substantive theory of truth. *Synthese*. **117**(1), pp.133-172.
- Sher, G. 2004. In search of a substantive theory of truth. *Journal of Philosophy*. **101**(1), pp.5-36.
- Sher, G. 2005. Functional pluralism. *Philosophical Books*. **46**(4), pp.311-330.
- Sher, G. 2013. Forms of correspondence: the intricate route from thought to reality. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.157-179.
- Sher, G. 2015. Truth as composite correspondence. In: Achourioti, T., Galinon, H., Martínez Fernández, J., and Fujimoto, K. eds. *Unifying the philosophy of truth*. Dordrecht: Springer, pp.191-210.
- Silk, A. 2015. How to be an ethical expressivist. *Philosophy and Phenomenological Research*. **91**(1), pp.47-81.
- Simmons, K. 1999. Deflationary truth and the liar. *Journal of Philosophical Logic*. **28**(5), pp.455-488.
- Smith, M. 1994a. Minimalism, truth-aptitude and belief. *Analysis*. **54**(1), pp.21-26.
- Smith, M. 1994b. Why expressivists about value should love minimalism about truth. *Analysis*. **54**(1), pp.1-11.
- Smith, M. 2002. Evaluation, uncertainty and motivation. *Ethical Theory and Moral Practice*. **5**(3), pp.305-320.
- Soames, S. 2010. *What is meaning?* Princeton, NJ: Princeton University Press.
- Stampe, D. 1977. Towards a causal theory of linguistic representation. *Midwest Studies in Philosophy*. **2**(1), pp.42-63.
- Steglich-Petersen, A. 2006. No norm needed: on the aim of belief. *Philosophical Quarterly*. **56**(225), pp.499-516.
- Stewart-Wallace, A. 2016. An old solution to the problem of mixed atomics. *Acta Analytica*. **31**(4), pp.363-372.
- Stoljar, D. 1993. Emotivism and truth conditions. *Philosophical Studies*. **70**(1), pp.81-101.
- Strawson, P. 1949. Truth. *Analysis*. **9**(6), pp.83-97.
- Strawson, P. 1950. Truth. *Proceedings of the Aristotelian Society*. **24**(1), pp.129-156.

- Strollo, A. [Forthcoming]. A simple notion of validity for alethic pluralism. *Synthese*. [Online]. [Accessed 2 December 2016]. Available from: <https://doi.org/10.1007/s11229-016-1280-0>
- Tappolet, C. 1997. Mixed inferences: a problem for pluralism about truth predicates. *Analysis*. **57**(3), pp.209-210.
- Tappolet, C. 2000. Truth pluralism and many-valued logics: a reply to Beall. *Philosophical Quarterly*. **50**(200), pp.382-385.
- Tappolet, C. 2010. Review of Michael P. Lynch, *Truth as one and many*. *Mind*. **119**(476), pp.1193-1198.
- Tarski, A. 1933. The concept of truth in formalized languages. In: Tarski, A. *Logic, semantics, metamathematics: papers from 1923 to 1938*. Oxford: Oxford University Press, pp.152-278.
- Tarski, A. 1944. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*. **4**(3), pp.341-376.
- Thomasson, A. 2016. Metaphysical disputes and metalinguistic negotiation. *Analytic Philosophy*. **57**(4), pp.1-28.
- Unwin, N. 1999. Quasi-realism, negation and the Frege-Geach Problem. *Philosophical Quarterly*. **50**(196), pp.337-352.
- Unwin, N. 2001. Norms and negation: a problem for Gibbard's logic. *Philosophical Quarterly*. **51**(202), pp.60-75.
- Väyrynen, P. 2013. *The lewd, the rude and the nasty: a study of thick concepts in ethics*. Oxford: Oxford University Press.
- Väyrynen, P. 2016. Thick ethical concepts. In: Zalta, E. ed. *The Stanford encyclopedia of philosophy (winter 2016 edition)*. URL = <<https://plato.stanford.edu/archives/win2016/entries/thick-ethical-concepts/>>.
- Velleman, D. 2000. *The possibility of practical reason*. Oxford: Oxford University Press.
- Vision, G. 2004. *Veritas: the correspondence theory and its critics*. Cambridge, MA: The MIT Press.
- Walker, R. 2001. The coherence theory. In: Lynch, M. ed. *The nature of truth: classic and contemporary perspectives*. Cambridge, MA: The MIT Press, pp.123-158.
- Wedgwood, R. 2002. The aim of belief. *Philosophical Perspectives*. **36**(s16), pp.267-297.
- Williams, B. 1974. *Problems of the self*. Cambridge: Cambridge University Press.
- Williams, B. 1985. *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.
- Williams, C. 1992. *Being, identity, and truth*. Oxford: Oxford University Press.
- Williams, J. 2008. Ontic vagueness and metaphysical indeterminacy. *Philosophy Compass*. **3**(4), pp.763-788.
- Williamson, T. 1994a. A critical study of Crispin Wright, *Truth and objectivity*. *International Journal of Philosophical Studies*. **30**(1), pp.130-144.
- Williamson, T. 1994b. *Vagueness*. London: Routledge.
- Williamson, T. 2000. *Knowledge and its limits*. Oxford: Oxford University Press.

- Wiredu, K. 2004. Truth and an African language. In: Brown, L. ed. *African philosophy: new and traditional perspectives*. Oxford: Oxford University Press, pp.35-50.
- Wittgenstein, L. 2009. *Philosophical investigations*. 4th ed. Hacker, P. and Schulte, J. eds. Oxford: Wiley-Blackwell.
- Woods, J. 2014. Expressivism and Moore's paradox. *Philosophers' Imprint*. **14**(5), pp.1-12.
- Wrenn, C. 2011. Practical success and the nature of truth. *Synthese*. **181**(3), pp.451-470.
- Wright, C. D. 2005. On the functionalisation of pluralist approaches to truth. *Synthese*. **145**(1), pp.1-28.
- Wright, C. D. 2010. Truth, Ramsification, and the pluralist's revenge. *Australasian Journal of Philosophy*. **88**(2), pp.265-283.
- Wright, C. D. 2012. Is pluralism about truth inherently unstable? *Philosophical Studies*. **159**(1), pp.89-105.
- Wright, C. D. and Pedersen, N. 2010. Truth, pluralism, monism, correspondence. In: Wright, C. D. and Pedersen, N. eds. *New waves in truth*. Basingstoke: Palgrave Macmillan, pp.205-217.
- Wright, C. J. G. 1992. *Truth and objectivity*. Cambridge, MA: Harvard University Press.
- Wright, C. J. G. 1993. *Realism, meaning and truth*. 2nd ed. Oxford: Basil Blackwell.
- Wright, C. J. G. 1994. Response to Jackson. *Philosophical Books*. **35**(3), pp.169-175.
- Wright, C. J. G. 1996. Précis of *Truth and objectivity*. *Philosophy and Phenomenological Research*. **56**(4), pp.863-868.
- Wright, C. J. G. 1998a. Comrades against quietism: reply to Simon Blackburn on *Truth and objectivity*. *Mind*. **107**(425), pp.183-203.
- Wright, C. J. G. 1998b. Truth: a traditional debate reviewed. *Canadian Journal of Philosophy*. **28**(sup1), pp.31-74.
- Wright, C. J. G. 2001. Minimalism, deflationism, pragmatism, pluralism. In: Lynch, M. ed. *The nature of truth: classic and contemporary perspectives*. Cambridge, MA: The MIT Press, pp.751-787.
- Wright, C. J. G. 2013. A plurality of pluralisms. In: Pedersen, N. and Wright, C. D. eds. *Truth and pluralism: current debates*. Oxford: Oxford University Press, pp.123-153.
- Wyatt, J. 2013. Domains, plural truth, and mixed atomic propositions. *Philosophical Studies*. **166**(1), pp.225-236.
- Yu, A. 2017. Can alethic pluralists maintain compositionality? *Philosophical Quarterly*. **67**(268), pp.625-632.
- Zangwill, N. 2005. The normativity of the mental. *Philosophical Explorations*. **8**(1), pp.1-19.