

Two Worlds, One Mind:
The Divide between Perception and Belief

By
Grace Helton

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Philosophy
New York University
May, 2015

Ned Block

© Grace Helton

All Rights Reserved, 2015

DEDICATION

For my parents, David and Mary Jane, for Charity, and for Adam.

ACKNOWLEDGEMENTS

This thesis has had an extraordinary number of benefactors, foremost among them my advisor, Ned Block, and my committee members, David Chalmers and Michael Strevens, who patiently helped me craft my arguments and refine my writing through several difficult stages. I am also grateful to Jim Pryor for providing me encouragement and philosophical guidance at a critical point early in my graduate career. To Susanna Siegel I owe an enormous intellectual debt, as her work played an important role in piquing my interest in philosophy of perception in general and in some of the particular issues treated in this thesis.

For helpful discussions of various aspects of the project, I would like to thank Tom Avery, David Bennett, Rosa Cao, Jeremy Dolan, Daniela Dover, Michelle Dyke, Nemira Gasiunas, Rob Hopkins, Zoe Jenkin, Alex Kiefer, Andrew Lee, Phillippe Lusson, Jack Lyons, Eric Mandelbaum, Jorge Morales, Michael Murez, Jake Quilty-Dunn, Adriana Renero, Gina Rini, Adam Schneit, Robert Schroer, Henry Shevlin, Susanna Siegel, Joulia Smortchkova, Pietro Snider, Amia Srinivasan, and Martin Zavaleta. The dissertation has also benefited from generous audience feedback at the 2013 APA Pacific Division Meeting, the CUNY Graduate Center Cognitive Science Speaker Series, Doc'in Nicod at Institut Jean Nicod, a colloquium at Hobart & William Smith Colleges, the NYU Consciousness Discussion Group, and several sessions of NYU Thesis Prep.

Finally, for friendship and emotional support during the particularly trying final stage of writing, I owe a special debt to Tom Avery, Jeremy Dolan, Charity Helton, Amanda MacAskill, Janelle Reichman, Adam Schneit, Jacob Shwartz-Lucas, and Marie Toruno.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
INTRODUCTION	1
1 The Revisability View of Belief & The Perception-Belief Divide	2
2 High-Level Perception	4
3 The Dissertation's Arguments	5
THE REVISABILITY VIEW OF BELIEF	11
0 Introduction	13
1 The Proposal: The Revisability View of Belief	15
1.1 The Revisability View and Masks	21
1.2 The Revisability View and Other Views of Belief	22
2 The Argument from the Norm of Revision	24
2.1 The Norm of Revision	25
2.2 Epistemic ' <i>Ought</i> ' Implies ' <i>Can</i> '	28
2.3 The Argument Completed	31
3 The Predictions of the Revisability View	33
3.1 States Permitted into the Class of Belief	33
3.2 States Excluded from the Class of Belief	35
3.3 Sincere Assertion, Motivational Role, and Inferential Promiscuity	38
4 Conclusion	41
HOW NOT TO ARGUE FOR HIGH-LEVEL PERCEPTION	43
0 Introduction	44

1 What Kinds of Features Can We See?	47
2 Introspective Strategies	51
2.1 Introspection and High-level Perception	51
2.2 The Veridicality Method: Tye's Argument for Particularism	52
3 A Limitation on Introspection: Attitudinal Opacity	54
3.1 Attitudinal Opacity	54
3.2 Criticizing Introspective Strategies	60
4 What Felt Unity Doesn't Show: Reply to Siegel	61
4.1 Siegel on Felt Unity	62
4.2 Felt Unity from Narrative Coherence	64
5 What Agnosia Doesn't Show: Reply to Bayne	67
5.1 Bayne's Argument for Perception of Artifactual Features	68
5.2 Associative Agnosia As a Deficit of Holistic Processing	69
6 Conclusion	72
VISUALLY PERCEIVING THE INTENTIONS OF OTHERS	74
0 Introduction	77
1 The Thesis and Why It Matters	78
2 The Initial Plausibility of the Visual Explanation	80
3 The Initial Plausibility of the Doxastic Explanation	84
4 Against the Argument from Judgment-discordance	86
5 Against the Doxastic Explanation of the 'Fleeing' Experience	93
5.1 The Revisability View of Belief	93
5.2 The Argument from Non-revisability	96
6 Is The 'Fleeing' Experience a System 1 Output?	101

6.1 The First Horn of the Dilemma:	
System 1 Outputs as Rationally Revisable	103
6.2 The Second Horn of the Dilemma:	
System 1 Outputs as Non-Consciously Persisting	107
7 Is the 'Fleeing' Experience a Delusion-Like Cognitive State?	109
7.1 Similarities between Delusions and the 'Fleeing' Experience	111
7.2 Against the Delusion-Based Explanation	113
8 Visually Perceiving the Intentions of Others	117
BIBLIOGRAPHY	120

Introduction

On one traditional picture of the relation between perception and belief, perception might be personified as the less intellectually sophisticated helpmate of belief. On such a picture, perception's primary job is to hand over to belief information that it cannot itself make sense of, and it is belief's job to interpret this information and make it available for further inference and action-guiding. What perception lacks in intellectual sophistication, it makes up for in a kind of rational innocence: perception serves as a direct conduit to features of the environment, in that the purely sensory, non-conceptual information it hands over to belief is uncorrupted by the background assumptions that can bias belief.

On the picture described, perception and belief are very different kinds of entities. They are distinguished by whether they can be influenced by background mental states, and by what kinds of features they can represent. A number of developments in contemporary philosophy of mind work together to upend this traditional picture and in so doing, challenge the presumption that the divide between perception and belief is a theoretically important one.

On the perception side of things, it is now widely (though not universally) held that perception: is underwritten by inference-like sub-personal processes; can be shaped by subject desires, expectations, and beliefs; and can represent a wide range of conceptually sophisticated features—such as *being a mandolin* or *being the cause of some event* (Bayne, 2009; Clark, 2013; Hohwy, 2013; Macpherson, 2012; Shea, 2014; Siegel, 2005; Siegel, 2010). On the belief side of things, it is now widely (though not universally) held that belief: is at least sometimes formed by non-inferential processes, and can persist in the face of evidence that conflicts with it (Byrne, 2009, pp. 450-1; Evans, 2008; Gertler, 2011; Huddleston, 2012; Mandelbaum, 2014; Mylopoulos,

2015, p. 768). Partly motivated by the erosion of these traditional markers of the boundary between perception and belief, some theorists have suggested that we reject altogether the view that perception and belief are of different fundamental kinds, and replace it with the view that perception and belief merely occupy different positions on a continuum of relevant gradable features (Clark, 2013, p. 10; Hohwy, 2013, pp. 68-69; Shea, 2014).

In this dissertation, I reaffirm one aspect of the traditional picture, by arguing that belief and perception can be distinguished by their respective rational roles. Partly relying on this proposed rational difference between perception and belief, I reject a different aspect of the traditional picture, on which perception cannot represent conceptually sophisticated features. Focusing on the visual modality, I argue that visual experience can represent at least some features other than shape, color, and movement. On the picture that emerges: (i) at least some mental states are both determinately perceptual and determinately not beliefs, and (ii) at least some of these states represent conceptually sophisticated features. This picture is consistent with the possibility that some mental states are in some sense 'between' perception and belief (Carey, 2009; Shea, 2014). But for at least some mental states, the perception-belief divide holds firm.

1 The Revisability View of Belief & The Perception-Belief Divide

I argue that belief and perception can be distinguished by their respective rational roles. Specifically, I argue that all beliefs, insofar as they are beliefs, are nomically capable of being rationally revised in response to any bit of available, sufficiently strong evidence that conflicts with them (for the purposes of this introduction, I will refer to this trait as *rational revisability* or just *revisability*). In contrast, at least some perceptual states lack this capacity. The motivation for the

view that beliefs are rationally revisable derives largely from philosophical—and more particularly, epistemological—considerations.¹ In contrast, the motivation for the view that at least some perceptual states are not rationally revisable derives from broadly empirical considerations.

The argument in favor of the view that beliefs are necessarily revisable rests on two premises. Very roughly stated, these are: First, all beliefs are such that they ought to be rationally revised in response to any bit of available, sufficiently strong evidence that conflicts with them. Second, any mental state that ought to be rationally revised is nomically capable of being rationally revised. These claims are both of a broadly epistemological stripe, in that the first posits a certain rational requirement on belief, and the second is an epistemic variant of the principle that ‘ought’ implies ‘can.’

The claim that at least some perceptual states are not rationally revisable is motivated by empirical considerations, such as the following: subjects who view the Müller-Lyer stimulus under normal conditions represent the two lines as different in length, even when they appreciate the lines to be of the same length (Schlottman & Shanks, 1992). Further, the illusory representation of the Müller-Lyer stimulus is persistent over time, as evidenced by the fact that even subjects who have seen the stimulus countless times—such as theorists of perception—continue to represent its lines as different in length.² Likewise, subjects who view the Heider and Simmel movie—

¹ This is to exaggerate the point a bit, as the view is also constrained by the capacities had by most putative beliefs in humans.

² These claims are meant to be consistent with evidence that the illusory interpretation of the Müller-Lyer stimulus does not arise for subjects who were raised in an environment whose architectural elements are less likely to include sharp corners (Segall, Campbell, & Herskovit, 1966). That environmental influences should contribute to the fine-tuning of the visual system during a sensitive period is not surprising, and it is independent of the issue of whether the visual experience of a developmentally mature organism is rationally susceptible to occurrent cognitive states.

which is an animated film depicting simple shapes in movement—represent the presented figures as animate and as acting on intentions, even though they appreciate that the figures are neither animate nor acting on intentions (Heider & Simmel, 1944; Hashimoto, 1966; Scholl & Gao, 2013). This ‘animacy’ representation of the movie is further persistent across time and cultures (Barrett et al., 2005; Scholl & Gao, 2013).

The view that at least some perceptual states are not rationally revisable should be carefully distinguished from the view that perception is *cognitively impenetrable*. Some state is cognitively impenetrable just in case that state cannot be rationally influenced in a certain way by *any* personal-level cognitive state (Macpherson, 2012; Pylyshyn, 1999).³ In contrast, some state is non-revisable just in case it is rationally immune to *at least some* cognitive states.

The claim that some perceptual state is cognitively impenetrable is stronger than the claim that that perceptual state is not rationally revisable. If some state is cognitively impenetrable, it is rationally immune to all other states, so it is also non-revisable. In contrast, some state might count as non-revisable in virtue of being rationally immune to merely a proper sub-set of all cognitive states, and such a state will not count as cognitively impenetrable. Thus, we can claim that some perceptual state is not rationally revisable without committing ourselves to the stronger claim that that state is cognitively impenetrable.

2 High-level Perception

By distinguishing between belief and (some kinds of) perception by appealing to different rational roles, we at least partly reaffirm one aspect of the traditional view of the perception-belief divide, on which belief is responsive to evidence in a way that perception is not. It turns out to be

³ Pylyshyn-style cognitive impenetrability permits that states might be influenced by cognitive states via certain indirect mechanisms, such as by diachronic intra-system changes or via a shift in the organism’s distribution of attention.

possible to use this proposed rational difference between belief and perception to argue against a different aspect of the traditional view, on which conceptually sophisticated features can be represented by belief but cannot be represented by perception.

In the contemporary literature, the view that we perceive certain *high-level* features has been defended by a range of theorists (Audi, 2013; Bayne, 2009; Bayne, 2011; Block, 2014; Butterfill, 2009; Chappell, 2008; Cullison, 2010; Farennikova, 2013; Fish, 2012; Nanay, 2011; Scholl & Gao, 2013; Scholl & Tremoulet, 2000; Siegel, 2005; Siegel, 2009; Siegel, 2010; Sorenson, 2012; van Gulick, 1994). In the visual domain, high-level features are any features other than shape, color, contrast, motion, or other paradigmatically *low-level* visual features. Despite the proliferation of arguments in favor of high-level perception, a persistent opposition to at least some kinds of high-level perception remains (Byrne, 2009; Chalmers, 2006; Pautz, 2009; Prinz, 2013; Rips, 2011). A third position has also emerged, on which there is no fact of the matter about whether we perceive certain high-level features (Logue, 2012).

Dissatisfied with extant arguments in favor of high-level perception, I present a different argument in favor of (some kinds of) high-level perception. In the first step of the argument, I identify certain target mental states and argue that these states represent high-level features and are not rationally revisable. I then claim that the fact that these mental states are not rationally revisable is best explained if we posit that these states are perceptual states.

3 The Dissertation's Arguments

The dissertation is structured as three thematically connected but self-standing essays. In the first essay, "The Revisability View of Belief," I argue in favor of *the revisability view of belief*, on which all beliefs are nomically capable of being rationally revised in response to any bit of sufficiently strong, available evidence that conflicts with them. As mentioned, the motivations for this view are broadly normative in character, but the view itself is a descriptive view. It is not a

view about what beliefs ought to do. It is rather the view that if some mental state is not revisable in the relevant sense, it is not a belief, though it may be some other kind of cognitive attitude, such as an entertained thought, a pretense, or a non-doxastic delusion.

As against the worry that the revisability view is too strong, in that it would excessively narrow the class of belief, I argue that the view permits into the class of belief a wide range of irrational states, including states sustained by confirmation bias and at least some emotionally-underpinned states. I further argue that in excluding other states from the class of belief—such as certain faith-based religious acceptances—the revisability view makes the right prediction. I argue for this in two steps: First, I claim that the intuition *states such as faith-based religious acceptances are beliefs* is rooted in the view that states that are sincerely asserted, action-guiding, and inferentially promiscuous are beliefs. Second, I argue that none of these conditions is sufficient for belief. I conclude that there is no principled reason to require that states such as faith-based religious acceptances be permitted into the class of belief. Further, the philosophical motivations that underpin the revisability view count as a strong reason to exclude such states from the class of belief.

In the second essay, “How Not to Argue for High-level Perception,” I turn to the question of whether perception can represent high-level features. I criticize several arguments in the literature aimed at establishing that we visually perceive high-level features.

I first consider introspective methods and argue that these offer no evidence whatsoever that might support the view that we perceive high-level features. For instance, introspecting on the experience of viewing a mandolin provides no evidence whatsoever in favor of the view that one visually perceives the object before one *as* a mandolin. To argue for this, I first suggest that the most plausible cognitive explanation of the disputed mental states will not appeal to slow, effortful, analytically-formed beliefs, but will rather appeal to *System 1 outputs*, which are fast,

automatic, heuristically-produced thoughts (Evans, 2008). On the System 1 output explanation of these states, these states are a combination of low-level perceptual experiences combined with a relevant System 1 assessment. For instance, when viewing a mandolin, one visually perceives shape, form, and color features and the perception of these features together elicit a System 1 assessment ‘that is a mandolin.’ The feature of being a mandolin is not represented in vision itself. Once we recast the dispute over high-level perception as a dispute over whether certain mental states are wholly perceptual or whether they are rather comprised of combinations of low-level perception and a relevant System 1 thought, it becomes implausible that introspection might adjudicate between these explanations.

I next consider a suggestion from Susanna Siegel that certain mental states that represent high-level features exhibit a kind of felt unity between visually perceived low-level features and represented high-level features and that this felt unity is explicable if the relevant high-level features are also visually perceived but is harder to explain if the high-level features are represented merely at the level of thought. For instance, viewing a ball landing on the ground at the same time that some houselights go off can elicit an illusory feeling of causation between the ball’s landing and the houselights’ going off. Siegel argues that in this case, the visual experience as of shape, motion, and color is felt to be unified with the representation as of causation and that this is best explained if the representation as of causation occurs at the level of visual experience (Siegel, 2009; Siegel, 2010). Against Siegel’s claim, I argue that in other cases, combinations of distinct attitudes can together elicit the relevant feeling of unity when those attitudes are near-simultaneous in onset and narratively unified. This suggests that a perception-thought combination might likewise elicit a feeling of unity in the ‘ball’ case and thus, that we need not posit visual perception as of causation to account for the felt unity in that case.

Finally, I consider an argument from Tim Bayne that the condition known as *associative visual form agnosia* is best explained by positing that typical subjects visually perceive certain high-level features. This form of agnosia involves a deficit in the ability to categorize visually-presented artifacts, combined with intact low-level visual perception. Bayne argues that the agnosic lacks visual perception as of artifactual features, such as *being a pipe* or *being a stethoscope* and that in this way agnosics differ from their typical counterparts, who enjoy visual perception as of artifactual features (Bayne, 2009).

Against Bayne's position, I adduce empirical evidence that the relevant form of agnosia is best explained as a deficit in the patient's ability to simultaneously visually integrate multiple spatial features (I draw in particular from Delvenne et al., 2004 and Farah, 1990). This form of agnosia is thus not a deficit in high-level perception *per se*. For this reason, agnosia does not suggest a reason to think typical subjects enjoy artifactual perception. Rather, what distinguishes agnosics from their typical counterparts is merely that agnosics cannot holistically visually perceive an object's spatial features, whereas typical subjects can.

In the third and final essay, "Visually Perceiving the Intentions of Others," I turn to outlining my positive argument in favor of (at least some kinds of) high-level perception, by using the experience as of others' intentions as a test case. I first argue that the typical subject viewing the Heider and Simmel movie represents the figures in the movie as acting on certain intentions. I then draw on empirical evidence to argue that this representation as of others' intentions is not rationally revisable. I then consider and reject three plausible cognitive explanations of this representation and claim that all three of these explanations struggle to offer a satisfying account of this state's unrevisability. Finally, I claim that by positing that this representation is a visual experience, we can satisfyingly account for the state's unrevisability.

On the first cognitive explanation I consider, the representation as of others' intentions is a post-perceptual judgment or belief. I reject this view on the grounds that the representation is not rationally revisable, whereas, according to the revisability view of belief, all beliefs are necessarily rationally revisable.

On the second cognitive explanation I consider, the representation as of others' intentions is a System 1 output. Against this view, I first argue that there are two empirically plausible views of the relation between System 1 outputs and conflicting evidence: On the first, System 1 outputs can be rationally revised in response to conflicting evidence. On the second, System 1 outputs cannot be rationally revised in response to conflicting evidence, but System 1 outputs that are contravened by relevant evidence are invariably moved to the level of the non-conscious, where they ultimately degrade for reasons of storage limitations. I then suggest that on either of these views, the representation as of others' intentions is not a System 1 output. If on the one hand, System 1 outputs are rationally revisable, they are no better poised to explain the representation as of others' intentions than judgments (indeed, on this view System 1 outputs are likely a species of judgment). If on the other hand, System 1 outputs are not rationally revisable but are moved to the non-conscious in the face of contravening evidence, it is also not the case that the representation as of others' intentions is a System 1 output. For the representation as of others' intentions does not become non-conscious in response to conflicting evidence; it remains conscious so long as the relevant stimulus is present.

On the third cognitive explanation I consider, the representation as of others' intentions is a delusion-like cognitive attitude. I criticize this view on the grounds that the mechanisms that plausibly explain why clinical delusions are not rationally revisable are not operative in the typical subject viewing the Heider and Simmel movie. Thus, positing that the representation as of others'

intentions is a delusion-like cognitive attitude does not offer a convincing explanation of why the representation as of others' intentions is not rationally revisable.

I finally claim that if we posit that the representation as of others' intentions is a visual experience, and more particularly a visual illusion, we can explain its unrevisability in a satisfying way, by appealing to either of two plausible explanations. On the first explanation, the visual system enjoys the capacity to process the relevant conflicting cognitions, but it lacks access to those representations, which prevents its processing them (Fodor, 1983; Fodor, 1984; Fodor, 2001, p. 63). On the second explanation, the visual system is restricted in the kinds of information it can process, and it is this processing limitation that explains why vision is unaffected by the relevant conflicting cognitions (Barrett, 2006; Carruthers, 2004; Carruthers, 2006; Carruthers, 2011). Either of these explanations permits us to offer a satisfying account of why the representation as of others' intentions is unrevisable: This representation is a visual illusion, so the fact that it is not rationally revisable can be accounted for by appealing to the mechanisms that underlie visual illusions more generally.

The Revisability View of Belief

1

It is widely held that for some mental state to be a belief, it must be, in some sense or other, responsive to evidence (Adler, 2002; Currie & Ravenscroft, 2002; Gendler, 2008; Shah & Velleman, 2005; Velleman, 2000; cf. Bayne & Pacherie, 2005; Bortolotti, 2011).⁴ The claim that beliefs are in fact evidence-responsive is distinct from the normative claim that beliefs *ought* to respond to evidence. The descriptive claim says that if some mental state is not evidence-responsive in the appropriate way, it is not a belief, though it may be some other kind of cognitive attitude, such as an entertained thought, a pretense, or a non-doxastic delusion.

Though many theorists endorse the view that beliefs are necessarily evidence-responsive, this claim is rarely argued for.⁵ Instead, it is presented as an obvious conceptual truth or is simply presumed on the way to arguing for other claims (Adler, 2002; Currie & Ravenscroft, 2002; Egan, 2009; Gendler, 2008). The lack of a cogent defense of the view is particularly troubling in light of empirical evidence that many beliefs are formed in response to very poor evidence, or fail to be

⁴ Also, the *interpretative view of mind*—of the kind associated with Davidson (1984) and Dennett (1989)—entails that beliefs are necessarily evidence-responsive in some way, though it is unclear exactly how strong this evidence-responsiveness must be. For a discussion, see Döring (1990).

⁵ A notable exception is found in Shah and Velleman (2005), who suggest that in order to distinguish beliefs from other cognitive attitudes, we must posit that beliefs are necessarily evidence-responsive. For a discussion of this argument, see van Leeuwen (2009).

revised even when contravened by excellent evidence.⁶ If the view that beliefs are evidence-responsive requires that we exclude all such states from the class of belief, this may count as a reason to reject the view.

In this essay, I develop and defend a particular version of the view that beliefs are necessarily evidence-responsive. This is *the revisability view of belief*, which says that if some mental state is a belief, then that mental state must be nomically capable of being rationally revised in response to any bit of available, sufficiently strong evidence that conflicts with it. Since the revisability view does not require that beliefs be *formed* in response to evidence, but requires merely that existing beliefs can be revised in response to evidence, the view is compatible with evidence that beliefs are frequently formed in response to very poor evidence. The revisability view can also accommodate the fact that beliefs frequently are not revised in response to conflicting evidence. So long as such states have a certain capacity to be revised, they can count as beliefs.

That the revisability view can accommodate irrational beliefs shows merely that the view crosses a hurdle any view of belief must ascend. It does not suggest a positive reason to accept the view. The centerpiece of the paper is such an argument, *the argument from the norm of revision*. This argument moves from a claim about belief's susceptibility to a certain norm of rationality, to the conclusion that all beliefs are capable of being revised in response to conflicting evidence. The key to this transition is a certain epistemic version of the principle '*ought*' implies '*can*.' Painted in the broadest of strokes, the argument is as follows:

- (1) All beliefs are rationally required to be revised in response to conflicting evidence.

⁶ For evidence that beliefs are sometimes formed in response to poor or no evidence, see Mandelbaum (2014). For evidence that beliefs are sometimes maintained in the face of conflicting evidence, see Nickerson (1998).

- (2) If some mental state is rationally required to be revised in response to conflicting evidence, then that mental state *can* be revised in response to conflicting evidence.
- (3) All beliefs can be revised in response to conflicting evidence.

In §1, I develop the revisability view of belief. In §2, I present the main argument in favor of the revisability view. In §3, I consider what predictions the revisability view makes of particular mental states, including faith-based religious views. In §4, I conclude.

o Introduction

Before proceeding to my main arguments, I want to say something about the way I am conceiving of beliefs, and the method I am using to investigate them. I am presuming that if beliefs exist at all, they exist whether or not humans recognize them or regard them as existing. Beliefs are in this respect like atoms of calcium, Joshua trees, and wind currents. They are entities we must posit to explain some interesting range of empirical phenomena. Beliefs differ in this respect from both money and the Canadian province of Saskatchewan, which plausibly exist in virtue of something like implicit consent and authoritative decree, respectively. The metaphysical presumption that beliefs exist independently of human responses immediately rules out that we might merely decide or stipulate what beliefs are.

My strategy of isolating beliefs is primarily one of provisional reliance on a cluster of core claims typically associated with beliefs. On this strategy, both philosophical and ‘folk’ platitudes about beliefs can be useful in the initial stages of theorizing, but these platitudes should be treated as revisable in light of disconfirming evidence. There are other cases in which a cluster concept can help one pick out some entity even when the relevant cluster of features turns out not to obtain in that entity. For instance, in some contexts, the cluster concept *the man drinking champagne in the corner* can help one identify what is really a woman drinking vodka in the corner. Likewise, the typical cluster of claims associated with beliefs—that they are action-guiding, inferentially

promiscuous, rationally coherent, like *that*, and so on—might turn out to be useful in picking out beliefs, and thus in ultimately discovering the nature of beliefs, even if the total cluster of claims should turn out not to hold of beliefs.⁷

One reason I treat platitudes about beliefs as provisional is that I take it to be a near-datum that we have beliefs. Thus, I am presuming that *eliminativism about beliefs*, on which humans do not enjoy beliefs at all, is a highly implausible view, one which is more implausible than at least some radically revisionist views about the nature of belief. This means that there are at least some cases in which, given the choice between doing without belief in our theorizing about human psychology, or adopting a highly revisionist theory of belief, we should embrace revisionism. For instance, if it should turn out that we have been massively misled about the relation between belief and action, such that belief never guides action in the appropriate way, it may be that we should reject the view that beliefs are constitutively action-guiding, instead of concluding that humans do not enjoy beliefs.

Finally, a locutional note: I am using ‘belief’ to pick out a very broad range of states, including occurrent, merely dispositional, endorsed, and non-endorsed states. Thus, my ‘belief’ includes what are sometimes called *judgments*, where these are occurrent states that are not necessarily reflectively endorsed.⁸ ‘Belief’ further picks out both states that are produced by an automatic, non-conscious, fast, and heuristic process and states that are produced by an effortful, conscious, and analytic process. This liberal usage of ‘belief’ reflects the ambitions of the current project, which aims to identify what all such states have in common.

⁷ This general approach owes much to Mandelbaum (2014).

⁸ For a discussion of the difference between belief and judgment, see Cassam (2010).

1 The Proposal: The Revisability View of Belief

In this section, I lay out my positive proposal, the revisability view of belief. On this view, in order for some mental state to count as a belief, it must have a certain capacity to be revised in response to conflicting evidence:

THE REVISABILITY VIEW OF BELIEF: if some mental state is a belief, then it is nomically capable of being rationally revised in response to any piece of available, sufficiently strong evidence that conflicts with it.

Put slightly more formally, the revisability view says: for all x such that x is a belief, and for all y such that y is some piece of available, sufficiently strong evidence that conflicts with x : x is nomically capable of being rationally revised in response to y .

The revisability view is a descriptive claim, not a normative claim. It is not a claim about how beliefs *ought* to be; it's a claim about how beliefs must be insofar as they are beliefs. The revisability view says that if some state is not nomically capable of being rationally revised in response to any bit of available, sufficiently strong evidence that conflicts with it, it is not a belief, though it may be one of the other *cognitive attitudes*, such as an entertained thought, an assumption, or a pretense.⁹ Cognitive attitudes represent some state of affairs as obtaining, and in this way contrast with *conative attitudes*, such as wishes and desires, which treat some state of affairs as to be obtained (Shah & Velleman, 2005; Velleman, 2000).

Before proceeding to the major components of the revisability view, there are three aspects of the view worth highlighting at the outset: first, the capacity to be revised in response to *any*

⁹ Throughout, all references to revision should be understood to be references to *rational* revision, unless stated otherwise.

piece of evidence does not entail a capacity to be revised *simultaneously* in response to all pieces of evidence. It may be that some belief can be revised in response to evidence that *p* and can be revised in response to some other piece of evidence that *not-p* even though that belief cannot be revised simultaneously in response to both pieces of evidence. This restriction is consonant with how we think of other capacities: that Janelle can swim a mile and can play the *Rhapsody in Blue* clarinet solo does not suggest that Janelle can swim a mile *while* playing the *Rhapsody* solo. Second, the relevant capacity to be revised is not to do with how a state is *formed*; it is strictly to do with how an existing state responds to evidence. Hence, states which are formed in response to good evidence but which subsequently lack the capacity to be revised are not revisable in the relevant sense.

Third, whether some mental state is revisable depends on whether that state can be revised in response to conflicting evidence. This raises the question of what the view says about mental states that are *never* contravened by evidence. Perhaps those mental states that represent obvious necessary truths, such as the judgment *it's not the case that p and not-p*, are never contravened by evidence. Or, if God is all-knowing, then perhaps God's mental states are never contravened by evidence.¹⁰ If there are mental states that are never contravened by evidence, these states trivially satisfy the requirement of revisability in virtue of never failing to be revised in response to conflicting evidence. For this reason, the revisability view permits such states into the class of belief.

On the revisability view, beliefs must be (i) nomically capable of being (ii) rationally revised in response to (iii) available, conflicting evidence. I will discuss each of these components in turn. In the sense that is relevant to the revisability view, some mental state is *nomically capable* of being revised just in case, in at least some worlds where the relevant subject's

¹⁰ I am grateful to Tom Avery and Martin Zavaleta for pressing me on this issue.

psychological mechanisms are held fixed, that mental state is revised.¹¹ For mental states occurring in typical humans, the relevant worlds are those in which mechanisms of human psychology are held fixed. For mental states occurring in typical octopuses or machines or extra-terrestrials, the relevant worlds are those in which the typical mechanisms of octopus or machine or extra-terrestrial psychology are held fixed.¹²

Further, whether some mental state is revisable depends on whether that mental state can be revised in the very subject in whom it occurs. A mental state thus can't count as capable of being revised in virtue of the fact that it would be revised if it were to occur in a different subject.¹³

While the revisability view requires that beliefs be capable of being revised in response to conflicting evidence, the view is silent about the nature of the processes that mediate this revision and relatedly, about whether subjects are aware of or voluntarily bring about this revision. It is

¹¹ This method of analyzing capability is loosely based on Lehrer (1976).

¹² Arguably, subjects can survive changes in the kind of creature they are. There is an issue of how to characterize revisability in light of this fact. It's at least conceivable that a goldfish might be turned into an orangutan and manage to persist as the very same creature that it was before its transformation (it won't be the same *species*, but it might be the very same creature). If this possibility is a genuine one, then we need to further restrict the range of worlds that are relevant to revisability. Otherwise, a mental state held by a goldfish might count as revisable in virtue of the fact that it would be revised, were that goldfish transformed into an orangutan. This result would make the revisability view too weak to be of much interest. We can guard against this outcome by restricting the range of relevant possible worlds to those in which the relevant subject persists as *the very same kind of creature* that she is in the actual world.

¹³ On some views of the persistence conditions of mental states, it is not possible for the very same mental token to occur in two different subjects, even at different times. On such views, the proposed limitation on revisability will be harmlessly redundant.

thus consistent with the revisability view that belief revision should generally occur non-voluntarily, non-inferentially, or outside of the subject's awareness.

The second component of the revisability view is that beliefs must be capable of being *rationally revised* in response to conflicting evidence. Some mental state is rationally revised in the relevant sense only if that mental state is revised: (i) in response to evidence, (ii) in the right direction, (iii) and via a non-deviant route. I will briefly sketch each of these conditions in turn.

First, for some mental state to be rationally revised, it must be revised in response to evidence. If you judge that today is Wednesday and then, as the result of an unfortunate encounter with lightning, lose this judgment, your judgment has been revised, but not in response to evidence. This route to revision is thus non-rational.

Second, for some mental state to be rationally revised, it must be revised *in the right direction*. Which direction is the right direction is dictated by the evidence. For instance, if the best evidence suggests that not- p , then for a mental state as of p to be rationally revised, it must decrease in strength, disappear altogether, or be suspended.¹⁴ If this state should increase in strength, it is not rationally revised.¹⁵

¹⁴ I mention belief suspension separately since it might constitute a sui generis kind (Friedman, 2013).

¹⁵ One way a mental state can count as revisable is by decreasing in strength in response to conflicting evidence. In order for some mental state to decrease in strength without disappearing altogether, it must permit of degrees of strength. Not all revisable states will permit of degrees but we might think of those that do as permitting of degrees that can range from 0 up to and including 1 on the real numbers. Where these mental states are beliefs, these degrees are called *credences* and might be helpfully construed as a measure of subjective confidence (Strevens, 2012). I am supposing that there cannot be states which represent to degree 0 that p . When a belief actually reaches (as opposed to merely approaches) a credence of 0 , it ceases to exist.

Finally, for some mental state to be rationally revised, it must be revised via a non-deviant route. The task of distinguishing deviant from non-deviant routes to revision is a very large one. Indeed, it is one of the primary projects of contemporary epistemology. For our purposes, it is sufficient to distinguish deviant from non-deviant routes ostensively, by a pair of contrastive cases.

First, consider a case in which you believe that extra-terrestrials do not exist. You then read in a reliable newspaper that the government has captured one. On the basis of the report, you relinquish your belief that extra-terrestrials do not exist. This is a non-deviant path to revision. Compare this case to one in which you read in the newspaper that the government has captured an extra-terrestrial, and the shock causes you to fall out of your chair and bang your head. By sheer coincidence, the blow obliterates your belief that extra-terrestrials do not exist. In this case, your belief is revised in response to evidence, but the route to revision is deviant and hence, the revision does not count as a rational revision.

The final component of the revisability view is that beliefs must be capable of being revised in response to *available, sufficiently strong conflicting evidence*. If evidence is construed as states of affairs, then for those states to count as *available* for a subject, that subject must be aware of those states of affairs. Further, the mode or presentation under which the subject represents these states of affairs (if any) must be the same mode of presentation (if any) under which the relevant belief is described.

Here and throughout, the relevant notion of evidence is meant to be neutral between external and internal individuations of evidence. On some internalist conceptions of evidence, evidence necessarily consists of consciously accessible mental states. On such conceptions, the re-

quirement that conflicting evidence be available is harmlessly redundant. On externalist conceptions of evidence, evidence is at least partly comprised of states of affairs (Kelly, 2008). On such conceptions, the requirement that evidence be available is a substantive requirement.¹⁶

Evidence is *sufficiently strong* in the relevant sense just in case it is strong enough to trigger a rational requirement that the relevant belief be revised. The motivation for characterizing sufficient strength in this way derives from the argument in favor of the revisability view.

Finally, *conflicting evidence* comes in two basic varieties. For a belief that p , conflicting evidence can be evidence in favor of some proposition q , where q is inconsistent with p .¹⁷ Or, it can be evidence that undermines p itself. For instance, suppose a subject believes, on the basis of a visual experience as of a bison, that there is a bison in the distance. Reliable testimony that there are no bison in the area would count as evidence that conflicts with this belief, since it is evidence in favor of a proposition that is inconsistent with the proposition believed. Evidence that one's visual system is dramatically malfunctioning would also count as evidence that conflicts with the relevant belief, but for a different reason: it undermines the evidence which was the basis for that belief.

¹⁶ Going forward, I sometimes abbreviate 'sufficiently strong, available conflicting evidence' with 'sufficiently strong conflicting evidence' or just 'conflicting evidence.'

¹⁷ Two mental states are *inconsistent* in the relevant sense just in case the propositions they each represent cannot simultaneously obtain. The relevant kind of conflict is thus that of logical conflict. Pairs of mental states whose contents are transparently contradictions of each other ('Austria is in the E.U.' and 'Austria is not in the E.U.') will count as inconsistent, as will pairs of mental states whose contents conflict in a less transparent way ('Clark Kent cannot fly' and 'Superman can fly').

1.1 *The Revisability View and Masks*

The revisability view ascribes to beliefs a certain capacity, and capacities can be *masked*. For instance, if some glass is capable of breaking, then in at least some possible worlds, it does break. This is consistent with the fact that in many worlds, the glass won't break, even if struck with force. Even a very fragile glass won't break when struck if it is wrapped in soft, thick padding. But this does not mean that the padding renders the glass no longer capable of breaking. The padding merely obscures the glass' capacity to break (Bird, 1998; Johnston, 1992).¹⁸

For at least many beliefs in humans, the conditions which tend to facilitate rational belief revision are those in which all of the following hold: the belief and the evidence which contravenes it are both conscious and attended; the belief is not underpinned by strong emotion; and the subject enjoys available processing capacity. Correspondingly, for such beliefs, typical masks of the capacity to be revised will include conditions in which: the belief or the evidence which contravenes it are unattended or non-conscious; the belief is underpinned by strong emotion; or the subject lacks available processing capacity.

There are two reasons we should not take the preceding list of masks to exhaust possible masks of revisability. First, though these conditions describe some beliefs held by humans, it may be that there are other beliefs held by humans which are revised under rather different conditions than those described and hence, whose masking conditions are different than those described.

A second reason to take the list of masks as non-exhaustive is that the revisability view is not restricted to beliefs in humans; it extends to all beliefs, both actual and merely possible, whether those beliefs occur in non-human animals, artificially intelligent beings, or extra-terrestrials. It is at least conceptually possible that some of these creatures' beliefs are masked by very

¹⁸ This characterization of masks borrows from the dispositions literature, which is the context in which Bird and Johnston discuss masks.

different conditions than those which mask ordinary beliefs in humans. For instance, while attention is generally a requirement of belief revision in humans, it is at least conceptually possible that there exist creatures for whom *inattention* facilitates belief revision and hence, for whom attention masks belief revision.

1.2 The Revisability View and other Views of Belief

Since the revisability view posits merely a necessary and not a sufficient condition on belief, it is not a full characterization of belief. As it turns out, revisability is probably not sufficient for belief. One reason for thinking this is that it may be that for a mental state to count as a belief, that mental state must play some motivational or action-guiding role for its subject. On this view, if a subject believes tomatoes are vegetables, she must be disposed to say and do certain things, like assert that tomatoes are vegetables, place her tomatoes in the crisper drawer she reserves for vegetables, or increase her consumption of tomatoes as part of an attempt to increase her vegetable intake.

Another reason for thinking that revisability is not sufficient for belief is that it may be that all beliefs, insofar as they are beliefs, must play characteristic phenomenal roles. For instance, it may be that the belief that there is beer in the fridge must dispose one to experience surprise when opening the fridge and finding no beer in it (Schwitzgebel, 2002).

A final reason for thinking that revisability is not sufficient for belief is that it may be that in order for some mental state to count as a belief, that mental state must be *inferentially promiscuous*, or available as a premise across a wide range of inferences (Glüer & Wikforss, 2013; Mandelbaum, 2014). On this view, if one believes there is beer in the fridge, then one must be able to exploit this belief as a premise in further inferences. For instance, the belief that there is beer in the fridge might satisfy this requirement by contributing to inferences such as: *There is beer in*

the fridge. If there is beer in the fridge, I don't need to buy more beer. So, I don't need to buy more beer.

Notably, the revisability view is consistent with all of these proposed additional requirements on belief. For while the revisability view posits that revisability is necessary for belief, it does not further stipulate that this is the only necessary condition on belief.

Finally, the revisability view should be distinguished from the view that belief aims at truth.¹⁹ These views come apart in both directions. One might hold that beliefs are necessarily capable of being revised in response to conflicting evidence and simultaneously deny that belief aims at truth. For instance, one might maintain that what accounts for the fact that beliefs are capable of being revised is not that beliefs aim at truth but that belief (or perhaps the organism as a whole) aims at some other outcome, such as internal consistency or holding views that are close enough to the truth for practicable action. Conversely, one might endorse the view that belief aims at truth while denying that all beliefs have a capacity to be revised in response to evidence. On this view, beliefs that entirely lack a capacity to be revised in response to evidence might be viewed as highly deviant beliefs, but they will nevertheless have truth as an aim; such states will merely be ill-equipped to achieve this aim.

In this section, I have sketched the major tenets of the revisability view of belief. On the revisability view, all beliefs, insofar as they are beliefs, are nomically capable of being rationally revised in response to evidence that conflicts with them. If there are mental states which are never contravened by evidence, such as those which represent evident necessary truths, these mental states trivially satisfy the condition of revisability and hence, can count as beliefs. States which

¹⁹ For defenses of the view that belief has truth as an aim, see Shah and Velleman (2005) and Wedgwood (2002). For a defense of the view that belief's aim (if any) is knowledge, see McHugh (2011).

altogether lack the nomic capacity to be revised are not beliefs, though they may be some other cognitive attitude, such as a merely entertained thought, an assumption, or a cognitive pretense. In the next section, I turn to the argument in favor of the revisability view.

2 The Argument from the Norm of Revision

In this section, I present a positive argument in favor of the revisability view, *the argument from the norm of revision*. This argument extends in full generality to all doxastic states, whether occurrent or dispositional, attended or unattended, unconsidered or reflectively endorsed, conscious or non-conscious, compartmentalized from other states or integrated with other states, heuristically-produced or inferentially-produced.

The argument is named after its central premise, which states, roughly, that beliefs are rationally required to be revised in response to any bit of available, sufficiently strong evidence that contravenes them. Since this claim is normative, it cannot not by itself illuminate the descriptive nature of belief. But combining this claim with an epistemic version of the principle ‘*ought*’ implies ‘*can*’ yields a surprisingly powerful argument in favor of the revisability view. Here is the argument in full. Throughout, m is an arbitrarily selected mental state m , and S is the subject in whom m occurs:

- (1) If S 's belief m is contravened by available, sufficiently strong evidence, then S has a pro tanto obligation to rationally revise m in response to that evidence.
- (2) If S is pro tanto obligated to revise m in response to available, sufficiently strong evidence that contravenes it, then S is nomically capable of rationally revising m in response to that evidence.
- (3) If S is nomically capable of revising m in response to available, sufficiently strong evidence that contravenes it, then m is nomically capable of being rationally revised in response to that evidence.

(4) If S 's belief m is contravened by available, sufficiently strong evidence, then m is nomically capable of being revised in response to that evidence. i.e., the revisability view of belief is true.

The heavy lifters in the argument are the first two premises: (1) is the norm of revision, and (2) is an epistemic version of '*ought*' implies '*can*.' As we shall see, (3) is a truism or a near-truism. I turn now to defending the argument, focusing attention on the first two premises.

2.1 *The Norm of Revision*

The first premise of the argument just is the norm of revision. It says that subjects who enjoy beliefs that are contravened by available, sufficiently strong evidence have a pro tanto obligation to rationally revise that belief in response to that evidence. Here as before, a belief is *rationally* revised only if it shifts in the right direction in response to evidence and via a non-deviant route.

Pro tanto obligations are obligations that retain their force even when trumped by more pressing obligations (Scanlon, 1998, p. 50). In this way, they contrast with all-things-considered obligations. The following case is illustrative: you have promised to attend your friend's viola performance. He is performing one solo at the beginning of a longer concert. On the way to the show, you encounter a badly injured child who needs medical attention. If you stop and help the child, you will certainly miss your friend's performance. In this case, you are pro tanto morally required to keep your promise to your friend, at the same time that you have a more pressing moral requirement to ensure that the child receives medical care. Thus, your all-things-considered obligation is to assist the child, but this does not change the fact that you have a pro tanto requirement to attend your friend's concert. That requirement is simply trumped by another more pressing obligation.

The distinction between pro tanto and all-things-considered requirements also obtains in the epistemic domain. The reason the norm of revision is articulated in terms of a pro tanto and not an all-things-considered requirement is that the subject whose belief is contravened by evidence may also have other evidence which supports a different all-things-considered epistemic obligation. For instance, if one believes, on the basis of a visual experience, that there is a bison in the distance, and one also has very good evidence that one's visual system is malfunctioning and more particularly is causing visual bison hallucinations, one is at least pro tanto rationally required to revise one's belief. However, if one also has evidence in the form of reliable testimony that there is in fact a bison in the distance, it may be that one's all-things-considered epistemic obligation is to maintain one's belief. Nevertheless, the pro tanto requirement to revise the belief does not disappear in light of the more pressing epistemic obligation; it is simply trumped such that it is not, all-things-considered, what one ought to do.

It remains to argue for the norm of revision. I am taking (1) to reflect an intuitive, core feature of belief but much more importantly, to at least partly explain why it matters whether some state is a belief or not. Thus to reject (1) would be to both dissociate belief from norms of rationality in a peculiar way and, what's worse, to deprive the category of belief of its particular theoretical interest.

Consider that to reject (1) would be to accept that there exists some belief, held by some subject *S*, such that: *S* has available, sufficiently strong evidence that contravenes that belief, and yet *S* is not so much as pro tanto rationally required to revise that belief. Certainly, we can describe a case that satisfies these conditions, but that doesn't suggest that such a description reflects a genuine and not merely an epistemic possibility. For any such proposed case, one might reasonably doubt whether such a state is a belief and not merely an entertained thought, a pretense, an assumption, or some other attitude altogether.

In favor of the norm of revision is that it captures our pre-theoretic intuitions about a range of cases. If you believe there is fruit on your kitchen table and then, walking into the kitchen, see that the fruit bowl is empty, you should revise your belief. If you believe your child did not shoplift from a local convenience store and subsequently view surveillance footage showing your child doing just that, you should revise your belief. If you believe God exists and subsequently come to believe that the suffering that exists in the world is inconsistent with the existence of God, you should revise your belief.

In all these cases, the relevant obligation is an obligation of rationality. In some of these cases, obligations of morality or of prudence may recommend different courses of action. For instance, it may be that morality requires that you believe your child when she says that she did not shoplift, even though the surveillance footage says otherwise. The presence of such an overriding moral obligation, however, would not make it the case that you lose the pro tanto rational obligation to revise your belief; the moral obligation merely trumps it.

Importantly, the norm of revision extends to beliefs that are unattended, formed on the basis of perception, or non-conscious. Insofar as these states are beliefs, they ought to be rationally revised in response to sufficiently strong, available evidence that contravenes them. If you believe that the Bowery runs from east to west, and then come to examine a map of Manhattan, paying special attention to the Lower East Side, you should revise your belief. If you judge, on the basis of a perceptual experience, that two lines in a figure you are viewing are of different lengths, and then come to learn that the figure is illusory, you should revise your belief. If you believe implicitly, as the undetected effect of watching too many commercials, that drinking fruit juice is healthy, and then come to read about the ill effects of fruit juice on insulin response, you should revise your belief.

A final observation about (1) before moving to (2): the connection to norms of rationality is distinctive of belief, in that there are at least many other attitudes which do not exhibit it. For instance, suppose that some subject merely entertains the thought that she is the ruler of Sweden, perhaps to amuse herself during a particularly dry philosophy talk. Suppose also that she has excellent evidence that she is not the ruler of Sweden. There is no *rational* requirement that this subject revise her entertained thought. There may be other reasons, such as prudential reasons, for her to abandon the entertained thought, but rationally speaking, there is nothing amiss about it.

It would appear then, that the fact that beliefs are susceptible to the norm of revision suggests something special about the nature of belief. The tantalizing hope is that we might exploit belief's susceptibility to this norm to learn something substantial about belief's nature, something that distinguishes it from at least some other attitudes. By combining (1) with an epistemic version of '*ought*' implies '*can*', we can do just that.

2.2 Epistemic '*Ought*' Implies '*Can*'

I now turn to (2): if *S* is pro tanto rationally required to revise *m* in response to available, sufficiently strong contravening evidence, then *S* is nomically capable of rationally revising *m* in response to that evidence. (2) is an epistemic version of '*ought*' implies '*can*.' It says that if a subject ought to revise a belief, then she can revise that belief, in some sense of '*can*.' As we shall see shortly, the relevant sense of '*can*' is relatively weak, in that it does not require that the subject can voluntarily bring about the relevant revision.

The primary support for (2) is that it falls out of a more general claim about what agents can be rationally required to do, given their psychological limitations. In general, agents who are cognitively incapable of bringing about some state of affairs cannot be rationally required to bring about that state of affairs, even when their more cognitively capable counterparts might be so

required. For instance, a typical two-month old human infant cannot be rationally required to discriminate between an inaccurate depiction of the human body and an accurate depiction of the human body; in certain circumstances, typical adults might be so obligated.²⁰ Likewise, someone who reads English but not Mandarin cannot be rationally required to notice a glaring contradiction between a Mandarin text and its English translation; in certain circumstances, someone who reads both languages fluently might be so required.

The question arises: why aren't infants rationally required to appreciate inaccuracies in depictions of the human body, when their adult counterparts might be so required? And why aren't non-readers of Mandarin rationally required to make assessments on the basis of Mandarin-encoded information, when their Mandarin-fluent counter-parts might be so required? If we accept (2), and accept that rational requirements entail a corresponding psychological capacity, then we enjoy a straightforward and elegant answer to these questions: it is because rational requirements entail a correlative psychological capacity that infants can't be rationally required to recognize inaccuracies in the depiction of the human body and that non-readers of Mandarin cannot be rationally required to appreciate a glaring problem in a Mandarin text. Psychological capacity serves as a limit on rational requirement. Anyone who rejects this view must offer some explanation of why we don't hold infants to the same rational standards as adults and why we don't rationally require adults to incorporate information they lack the capacity to understand.

In the contemporary literature, epistemic versions of '*ought*' implies '*can*' have been widely discussed, but it is important to note that these disputes are for the most part orthogonal to whether we ought to accept the '*ought*' implies '*can*' of (2).²¹ This is because the contemporary

²⁰ Two-month-old infants cannot discriminate accurate from 'scrambled' depictions of the human body (Slaughter et al., 2011, pp. 87-91).

²¹ See Mizrahi (2012) and Ryan (2003) for recent criticisms of epistemic '*ought*' implies '*can*.' For

literature on epistemic versions of *'ought implies can'* is primarily concerned with versions of the principle that entail a voluntary capacity to revise one's beliefs. In contrast, (2) is neutral on whether the relevant subject can voluntarily bring about the relevant revision. Since our ultimate concern is not with the nature of the agent's capacity to revise her belief, but is rather with the revisability of the belief itself, it is sufficient for our purposes that the relevant belief revision be appropriately causally related to the agent or some of the agent's states, whether or not that causal relation is the product of voluntary action.²²

Since the kind of capacity relevant to (2) does not require a capacity for voluntary action, the *'ought implies can'* of (2) survives counter-examples which threaten stronger versions of *'ought implies can.'* For instance, consider the following case from Sharon Ryan, designed to refute one such stronger variant of *'ought implies can'*:

STICKY FINGERS

Your kleptomaniac friend, Sticky Fingers, has been accused of stealing your most prized possession. You are fond of Sticky Fingers and trust her completely. You believe she is innocent of the crime. After an investigation by the police, you are presented with conclusive evidence that Sticky Fingers committed the theft, but your belief in her innocence does not waver. You are simply psychologically incapable of voluntarily revising or relinquishing your belief in your friend's innocence. Nevertheless, so far as rationality is concerned, you ought to give up this belief (Ryan, 2003, p. 59).²³

recent defenses, see Hattiangadi (2010), Littlejohn (2012), and Vranas (2010).

²² For recent criticisms of belief's voluntariness, see Hieronymi (2006) and Ryan (2003). For a recent defense, see Steup (2008).

²³ This case is slightly modified from Ryan's so that it concerns already existing beliefs. Also, though Ryan's initial description of the case does not explicitly state that it is a voluntary capacity to revise one's beliefs that is relevant, the subsequent discussion of the passage makes explicit that it is a voluntary capacity that is at stake.

It may be that STICKY FINGERS succeeds as a counter-example to versions of *'ought' implies 'can'* that require that beliefs can be revised voluntarily. However, STICKY FINGERS does not rebut the version of *'ought' implies 'can'* outlined in (2) since this version does not require a voluntary capacity for revision.

Moreover, if the case were modified in such a way that would render it a threat to (2), we would no longer enjoy the intuition that you are rationally required to revise your view of Sticky Finger's innocence. Consider what such a case would involve: your view in your friend's innocence would have to be such that, holding fixed your psychological mechanisms, there is *no* possible world in which your view in your friend's innocence is rationally revised. Even if your warm feelings towards your friend were to disappear entirely, such that you no longer held her in any particular esteem, your view of her innocence would still not be revised in response to the evidence. Even if you were to spend years considering the evidence against your friend, your view would still not be revised.

It seems to me that in this case, in which your view of your friend's innocence is genuinely nomically incapable of being rationally revised, it is obscure in what sense you might be rationally required to revise your view. Perhaps it would be rationally *better* if you were to revise your view, but some outcome can be rationally better without being rationally required. Consider an analogy from the moral domain: the world would be a better place if you were to leap a mile-long chasm to rescue an imperiled child. However, you are not morally required to bring about this event. You simply can't leap a mile-long chasm, so you are not so obligated. Likewise, if your view of your friend's innocence can't be revised, you cannot be required to revise it.

2.3 The Argument Completed

The final premise in the argument is (3), which, to simplify somewhat, states: *If S is capable of revising m, m is revisable*. I am taking (3) to be a kind of truism. If someone is capable of

driving *that* car, that car is capable of being driven. If someone can cook *that* eggplant, that eggplant is capable of being cooked. For any particular, if someone can perform some action on that particular, then that particular is capable of having that action performed on it. Token beliefs fall under this general schema. If some subject is capable of bringing about a revision of some particular belief, then that particular belief is capable of being revised.

Finally, (4) is the conclusion, and states that for any arbitrarily selected belief, that belief is nomically capable of being rationally revised in response to available, sufficiently strong evidence that contravenes it. In other words, the revisability view of belief is true.

What exactly does this argument show? As it turns out, it tells us something quite surprising and informative about the nature of belief. Not only are beliefs capable of being revised in some bare metaphysical sense—even states very different from belief, like desires, emotions, and pretense might exhibit a capacity as weak as that—all beliefs, insofar as they *are* beliefs, are such that in at least some nomically possible worlds where they are contravened by evidence, they are revised.

I take the argument developed in this section to suggest a strong reason for positing that beliefs are necessarily nomically capable of being rationally revised. In effect, the argument says that any opponent of the revisability view must pay a cost: on the one hand, she might give up on the norm of revision, which would be to dissociate belief from rational requirement in a surprising way and in a way that would deprive belief of its theoretical interest. On the other hand, she might give up the view that a rational requirement entails a nomic capacity to be revised. But this would deprive her of a straightforward explanation of why subjects who lack a nomic capacity to bring about some state of affairs cannot be rationally required to bring about that state of affairs.

In the next section, I consider what predictions the revisability view makes of particular cases. I argue that mental states that are sustained by confirmation bias can count as beliefs, but

that faith-based religious views and similar states cannot. I further argue that the exclusion of faith-based religious views and similar states from the class of belief—though initially counter-intuitive—is ultimately a desirable result.

3 The Predictions of the Revisability View

Many beliefs held by actual humans are irrational. In developing a descriptive theory of belief, it is important to allow for this fact. If we don't, we risk ending up with a theory of good belief, where what we wanted was a theory of belief (Huddleston, 2012). In this section, I show that the revisability view is not at risk of ending up a theory of good belief. It permits at least some irrational states into the class of belief, including at least some states which are the result of confirmation bias and at least some states which are emotionally underpinned.

3.1 States Permitted into the Class of Belief

The phenomenon of confirmation bias occurs when subjects ignore or devalue evidence which conflicts with their existing views and attend to or overvalue evidence which supports their existing views. Confirmation bias is widespread in human reasoning; it has been observed in the context of paranormal beliefs, political beliefs, racist beliefs, and the pessimistic beliefs that are associated with certain anxiety disorders (Nickerson, 1998).

In a representative paradigm investigating confirmation bias, subjects were asked to indicate on a pre-defined scale the degree to which they were in favor of or against *neuro-enhancement*, or the non-palliative use of medicine for the purpose of improving cognitive, artistic, or athletic abilities. They were then given brief descriptions of eight different arguments, four of which were in favor of neuro-enhancement, and four of which were against it, and were asked to choose one of the eight arguments to read. Subjects who had self-rated as in favor of neuro-enhancement tended to select an argument in favor of neuro-enhancement. Subjects who had self-

rated as against neuro-enhancement tended to select an argument against neuro-enhancement. Thus, subjects avoided evidence that might challenge their pre-existing views. After reading their chosen argument, the subjects exhibited very little shift in their views (Schwind et al., 2012)

Arguably, mental states which persist due to confirmation bias are irrational. At least, the strategy which sustains them is at odds with the widespread assumption in philosophy of science that the best way to test a theory is to try to falsify it. Nevertheless, the revisability view can count at least some mental states that are sustained by confirmation bias as beliefs. The revisability view says that all beliefs must be nomically capable of being revised in response to evidence. And, as it turns out, at least some mental states which are sustained by confirmation bias are capable of being so revised. For instance, in a variant of the study of subjects' views of neuro-enhancement, subjects were encouraged to read an argument that was *inconsistent* with their reported view. This simple intervention resulted in subjects' moderating their initial views of neuro-enhancement (Schwind, et al., 2012). This demonstrates that in at least some cases, confirmation bias can be remediated simply by drawing subjects' attention to potentially disconfirming evidence; these states are thus nomically capable of being rationally revised.

Another class of beliefs which tend to resist evidence are emotionally underpinned beliefs. For instance, suppose that your child is accused of shoplifting cigarettes from a local convenience store. Suppose further that you enjoy evidence that your child in fact stole the cigarettes but that you nevertheless maintain her innocence. The question is: are you capable of rationally revising your view in response to the evidence? This depends on the particulars of the case.

There are practical and ethical problems in directly testing whether particular emotionally underpinned states would be revised if dissociated from emotion. Thus, it is difficult to say, of any particular emotionally supported state, whether that state is revisable. Nevertheless, there is indirect evidence that at least some such states—perhaps including your view that your child did not

steal the cigarettes—would be revised if divorced from feeling or if contravened by very strong evidence.

A primary mechanism of belief revision in humans involves *cognitive dissonance*, which is a kind of discomfort triggered when a subject experiences her views as in conflict. When subjects experience dissonance, they tend to revise one of their conflicting views in the direction of coherence with the other view. The dissonance itself—the feeling of discomfort—plays an essential role in this process (Harmon-Jones & Harmon-Jones, 2007; Elliot & Devine, 1994). This suggests that the catalyst of belief revision is the motivation to reduce dissonance. When the challenged belief *itself* is such that, giving it up would cause emotional distress, it may be preferable for a subject to remain in dissonance and to suffer the minor discomfort associated with it, then to suffer the greater emotional cost of giving up a dearly held belief.

Thus, at least some emotionally underpinned states—perhaps including your view that your child is innocent—may have a masked capacity to be rationally revised. Where the negative feeling associated with cognitive dissonance is not strong enough to ‘unmask’ that capacity, those states will remain unrevised. If your view in your child’s innocence is such a state, then the revisability view can admit it into the class of belief. If, on the other hand, your view in your child’s innocence lacks altogether a nomic capacity to be rationally revised, perhaps because it is constitutively tied to positive feelings about your child, then the revisability view will exclude it from the class of belief.

3.2 States Excluded from the Class of Belief

So far, I have shown that the revisability view permits at least some irrational states into the class of belief, including at least some mental states that are sustained by confirmation bias and at least some emotionally underpinned states. Whatever else we might say about the revisability view, it does not commit the error of winding up a theory of good belief.

Nevertheless, the revisability view makes certain predictions that may seem counter-intuitive. For it excludes from the class of belief any mental state which is not capable of being rationally revised, even if that mental state: guides action, is sincerely endorsed by its subject, and serves as a premise in a wide range of inferences. There are two kinds of mental states which satisfy this description: the first kind is comprised of states which are neither formed in response to evidence nor subsequently enjoy a capacity to be revised in response to evidence. It may be that some faith-based religious views are like this. The second kind is comprised of states which are formed in response to good evidence, but which altogether lack a capacity to be subsequently revised. It's unclear whether states of this second sort are common in humans, but they are at least conceptually possible; call such states *idées fixes*.

First, consider the subject who makes a Kierkegaardian leap of faith and accepts—in some sense of ‘accepts’—that God exists. Suppose that the resulting acceptance lacks any capacity to be revised in response to conflicting evidence. Suppose further that this acceptance plays a substantial role in motivating behavior—for instance, it explains why its subject engages in prayer, attends religious services, and the like—and that this acceptance also is sincerely and explicitly endorsed by its subjects. Finally, suppose this acceptance is inferentially promiscuous, in that it is available as a premise in a wide range of inferences, such as: *If God exists, we should love our neighbors. God exists. So, we should love our neighbors*. All of these features would seem to suggest that this acceptance is a belief, its unrevisability notwithstanding. However, if this acceptance is genuinely nomically incapable of being rationally revised, the revisability view excludes it from the class of belief.

Next, consider a subject who forms the view that her neighborhood’s farmer’s market takes place on Fridays. This view is formed in response to excellent evidence. Some point after forming this view, this subject suffers a minor brain lesion which leaves her cognitive faculties entirely

intact except for the curious result that she cannot revise her view about the local farmer's market. She sees flyers advertising that the market has been rescheduled for Sundays, her friends repeatedly tell her the market is now on Sundays, she even visits the farmer's market on Sundays (quite by accident, since she doesn't anticipate its being held then), but she simply cannot revise her view that the market is held on Fridays. She shows up at the usual spot every Friday, bags and shopping list in tow. She tells anyone who asks that the market is on Fridays. She relies on this claim as a premise in a wide range of inferences, such as: *The market is on Fridays. Today is Friday. The market is today.* The lesion has transformed her prosaic belief into an *idée fixe*.²⁴ On the revisability view, this mental state is not a belief, even though it was formed in response to good evidence. It may, however, be some other cognitive attitude, such as an entertained thought, an assumption, a cognitive pretense, or a non-doxastic delusion.²⁵

That the revisability view excludes faith-based religious acceptances and *idées fixes* from the class of belief might suggest that the view is too strong. After all, these states are action-guiding, sincerely endorsed by their subjects, and inferentially promiscuous. If they do not count as beliefs, we should like some explanation of why they do not. In the absence of such an explanation, it may seem that we should reject the revisability view, in favor of the following view:

THE ANTI-REVISABILITY VIEW OF BELIEF: At least some beliefs are not nomically capable of being revised in response to available, sufficiently strong evidence that conflicts with them.²⁶

²⁴ I am presuming a certain view of the persistence of mental states, on which a particular mental state can survive changes in the kind of attitude it is. This is not an essential part of the story, though. It could be that the subject's belief was destroyed and replaced by a non-doxastic delusion with the same content.

²⁵ Special thanks to Michael Strevens and Jeremy Dolan for pressing me on this sort of case and to Strevens for the helpful name.

²⁶ Something like the anti-revisability view is accepted by Gertler (2011), Huddleston (2012),

3.3 Sincere Assertion, Motivational Role, and Inferential Promiscuity

In this section, I wish to defuse the intuition that unrevisable faith-based religious acceptances and *idées fixes* are beliefs. I do this by first, suggesting that the source of this intuition is that these states have many of the typical traits of belief: they guide action, they are sincerely endorsed by their subjects, and they exhibit inferential promiscuity. Second, I argue that none of these traits is sufficient for belief and hence, the fact that faith-based religious acceptances and *idées fixes* exhibit these traits should not be taken to entail that these states are beliefs. Thus, I aim to defuse the intuition that these states are beliefs by undermining the model of belief that undergirds the intuition.

First, consider motivational role. It may be thought that if a mental state governs action in the right way—for instance, disposes one to attend church, pray, and attempt proselytization of others—that this is sufficient for its being a belief. However, as many theorists have by now pointed out, it is not only belief which can motivate action. Arguably, pretenses and suppositions also guide action (Gendler, 2007; Gendler, 2008; Velleman, 2000). For instance, suppose you are pretending to be an elephant. You might wave your trunk and walk clumsily and slowly. You don't (let's stipulate) believe you are an elephant. Plausibly, your pretense that you are an elephant itself motivates these actions (Velleman, 2000). It may be that there are certain kinds of motivational roles that only belief can play, but it is at least not obvious what these would be.

Second, consider sincere assertion. It may be that as a general rule, if some subject sincerely asserts *p*, that subject believes *p*. But there are cases in which sincere assertion that *p* occurs even when the subject does not believe *p*. The reason for this is quite simple: subjects don't always

Mandelbam (2014), Bayne and Pacherie (2005: 183), and Bortolotti (2011: 124).

know what they believe and can also have false beliefs about what they believe. For instance, consider a case of self-deception in which a subject believes that his husband is cheating on him but cannot admit this to himself. Such a subject might sincerely assert that his husband is faithful, while nevertheless exhibiting behavior that is consistent with his belief, such as feeling sad when his husband calls yet again to say that he will be working late, and asking his husband more questions than usual about his whereabouts. If this sort of case is so much as possible, sincere assertion that p does not entail belief that p .²⁷

Finally, consider inferential promiscuity, which is availability as a premise in a wide range of inferences. Mental states that are inferentially promiscuous have a kind of productive power; they can inferential-causally generate new mental states. Inferential promiscuity may be a necessary condition on belief, but it is not sufficient for belief. Consider that attitudes other than belief, such as supposition, also exhibit inferential promiscuity. For instance, suppositions can act as premises in arguments by *reductio*.²⁸

Consider the geometry student tasked with proving that no triangle has four sides. To do this, this student might suppose, for the sake of proving otherwise, that there is a triangle that has four sides and then attempt to generate a contradiction, as part of a demonstration that the original supposition is false. We might say that this student ‘hypothetically adds to her stock’ of beliefs that there is a triangle that has four sides, but whatever it is to hypothetically add to one’s stock

²⁷ Cohen (1992, pp. 68-73) also argues that sincere assertion does not entail belief.

²⁸ Strictly speaking, mental states do not serve as premises in arguments. Arguments are sets of propositions structured by a relation of putative validity. As such, they are abstract, mind-independent entities—certainly not the sorts of things that contain mental states. Saying that a mental state ‘serves as a premise’ in an argument is a loose way of saying that a mental state figures in a *psychological inference*. Psychological inferences are sets of mental states structured by a relation of putative validity; they are the mechanisms by which subjects grasp arguments.

of beliefs that there is a triangle that has four sides, it is not to *believe* that there is a triangle that has four sides. It is to suppose it or merely entertain it.²⁹

Given these considerations, we should be at least doubtful whether motivational role, sincere assertion, or inferential promiscuity are sufficient for belief. So, we should be at least doubtful whether faith-based religious views and *idées fixes* are beliefs. Thus, that the revisability view excludes such states from the class of belief does not suggest a reason to reject the revisability view.

Moreover, the argument in favor of the revisability view is simultaneously a reason for excluding faith-based religious views and *idées fixes* from the class of belief: because the norm of revision and ‘*ought*’ implies ‘*can*’ cannot be simultaneously true of these states, we should not class them as beliefs.

Consider again the subject who suffers from an unrevisable *idée fixe* that the farmer’s market is on Fridays. Notice that the following two claims—which are instances of the first two premises of the argument from the norm of revision—cannot be simultaneously true of this subject:

(1*) If *S*’s mental state that represents ‘the farmer’s market is on Fridays’ is a belief, then *S* has a pro tanto requirement of rationality to revise that mental state.

²⁹ Certain kinds of counterfactual reasoning provide more evidence of the inferential promiscuity of suppositions. Consider a case in which a consumer researcher asks you fill out a survey about food choices, which includes the following question: *Suppose you are having pizza for dinner. What would you order with it? Beer or wine?* One way to answer this question would be to simply comply with the task instructions, and to *suppose* that you are having pizza for dinner and then—by using inductive evidence from ‘seeing’ what your choice would be in that counterfactual situation—generate a hypothesis about what you would do in the relevant counterfactual. But of course, you don’t *believe* you are having pizza for dinner. You merely entertain the thought or suppose it, and this supposition plays an essential inferential role in the generation of your hypothesis.

(2*) If S has a pro tanto requirement of rationality to revise her mental state that represents ‘the farmer’s market is on Fridays,’ then S is nomically capable of so revising that mental state.

The subject’s *idée fixe* cannot simultaneously satisfy (1*) and (2*). But as the discussion in §2 demonstrated, we have independent grounds for accepting the broader principles of which (1*) and (2*) are merely instances. If the subject’s *idée fixe* really is a belief, it is subject to the norm of revision and hence, should satisfy (1*). But if it is subject to the norm of revision, it should enjoy a correlative nomic capacity to be revised and hence, should satisfy (2*). That the *idée fixe* cannot meet both of these requirements counts as a positive reason to exclude it from the class of belief.

In short, because *idées fixes* exhibit belief-like traits—such as inferential promiscuity, motivational role, and a connection to sincere assertion—we mistook them for beliefs, because many mental states with those features are beliefs. But closer reflection reveals *idées fixes* not to be beliefs, but rather to be pretenses, assumptions, or non-doxastic delusions. The same points apply *mutatis mutandis* to faith-based religious acceptances.

4 Conclusion

I have developed and defended the view that all beliefs are necessarily nomically capable of being revised in response to available, sufficiently strong contravening evidence. I have argued that this view is weak enough to accommodate beliefs that are for contingent reasons unresponsive to evidence. At the same time, the view is strong enough to accommodate belief’s susceptibility to the norm of revision. Views which reject a connection between belief and a nomic capacity to be revised struggle to accommodate belief’s susceptibility to this norm.

In the final essay in this dissertation, I put the revisability view of belief to work as part of an argument that we can visually perceive features other than shape, color, and movement. For instance, we sometimes visually perceive the intentions of others. Just as we can see something

as blue or as moving to the right, so too can we see someone as aiming to evade detection or as intending to traverse a physical obstacle.

How Not to Argue for High-level Perception

2

Suppose you view, under normal conditions, a person across from you. You experience—in some sense of ‘experience’—the object before you *as a face*. On one view, you *visually perceive* the object before you as a face, just as you visually perceive the object’s color and shape. On another view, you merely perceive certain color and form features, and on the basis of this visual experience, form the post-perceptual *thought* ‘that is a face.’ The claim that in this case you visually perceive the object as a face is an instance of the claim that we perceive *high-level* features. In this essay, I consider three strategies in the literature aimed at establishing that we perceive high-level features, such as *being a face*. I first consider arguments from several theorists that introspection provides *prima facie* evidence in favor of high-level perception. I argue that while introspection can provide evidence that high-level contents are *somehow* represented in experience, introspection does not suggest a reason to think that high-level contents are represented at the level of perception, and not merely at the level of a post-perceptual thought or judgment. I second consider an argument from Susanna Siegel that certain high-level experiences are felt to be unified with low-level visual information and that this felt unity is best explained by positing high-level visual perception. I argue that felt unity might equally be elicited by perception-thought combinations. I third consider an argument from Tim Bayne that certain forms of *agnosia* require that we posit that typical cognizers can visually perceive artifactual features, such as *being a guitar*. I marshal empirical evidence that the relevant form of agnosia is better-explained as a deficit in the simultaneous visual perception of multiple spatial features, and not as a deficit in artifactual perception *per se*.

The essay is structured as follows: In §1, I characterize high-level perception in further detail and explain some reasons why it matters whether we can perceive high-level features. In §2, I describe two broadly introspective strategies of arguing for high-level perception: these are the method of simple introspection and the method of appealing to veridicality judgments. In §3, I argue that neither of these introspective strategies succeeds, for the reason that introspection alone cannot rule out a plausible competitor view, on which fast, automatic, non-consciously processed post-perceptual thoughts account for the relevant phenomena. In §4, I describe and reject Siegel's argument that the fact that certain high-level experiences exhibit a felt unity suggests a reason to think these experiences are perceptual. In §5, I consider and criticize Bayne's argument that accounting for a certain form of agnosia requires that we posit that typical subjects visually perceive artifactual features. In §6, I summarize the key arguments and briefly sketch other extant strategies of arguing in favor of high-level perception.

o Introduction

Before proceeding to the main discussion, it may be helpful to clarify some framing issues. First, the present topic is perception, and I am relying on an ostensive understanding of this phenomenon. I take my starting point to be the paradigm cases: seeing blue, smelling lavender, tasting lemon,³⁰ and so on. Minimally, perception must originate in detection of features of the environment, but this requirement is meant to be consistent with the possibility of substantial processing at higher levels. Beyond the requirement of environmental feature detection, I am not supposing any particularly robust view about the characteristics of perception.

³⁰ Burge (2010) questions whether taste or smell are perceptual, since it is not clear that either exhibits object constancy. For present purposes this issue does not matter, since all of the cases discussed in this essay are to do with the visual modality.

Second, the issue of whether we can perceive high-level features in the relevant sense is the issue of whether we can perceive some entity *as* bearing a high-level feature, whether or not that entity in fact bears that feature. Thus, in the relevant sense, it might turn out that a subject visually perceives a particular object *as* a guitar even if the relevant object is in fact a mandolin which, due to viewing angle and special lighting effects, looks like a guitar.

The locution ‘perception *as of* Φ ’ is generally non-committal as to whether the relevant perceptual state is accurate, and the locution ‘perception *of* Φ ’ is sometimes reserved for accurate perception. I generally adhere to this convention, but occasionally abandon it for the sake of naturalness. Throughout, all talk of perception should be understood to be talk of ‘*as of*’ perception.

Third, I have already been talking as though perception represents features of the world, and thus that the objects of perception are contents.³¹ However, the issue of whether high-level features are perceived arises on both representational and non-representational views of perception’s objects. For those who view perception as a matter of subjects standing in the right relation to features of their environment, the question of whether high-level features are perceived collapses into the question of which features of the environment can stand in the right, perception-constituting relation to subjects. On one view, it is merely shape, color, and movement features which can stand in the visual relation to their subject. On another view, other features, such as being the cause of a certain effect can also stand in the visual relation to their subject. Though the issue of high-level perception arises equally for representational and for non-representational theories of perceptual objects, I will continue to talk in representational terms. Those inclined to a non-representational framework can reframe all of the present points in non-representational terms.

³¹ I have been and will continue to be neutral on the further issue of whether perceptual contents are Fregean intensions or are rather something more akin to Russellian states of affairs.

Another issue that is orthogonal to the issue of high-level perception is that of how perception is structured. On one view, perception is structured propositionally. On another view, perception is structured by the attributive schema ‘that F.’³² Importantly, whether high-level features are perceived is independent of the question of structure. For theorists who adopt the view that perception’s objects are modeled by the attributive schema ‘that F,’ the question of whether we can perceive (say) *being a face* becomes the question of whether we can enjoy an experience as of *that face*. While I sometimes talk as though the contents of perception are propositional, this does not reflect a deep assumption. Those inclined toward an attributive or other non-propositional model of structuring perceptual content are free to reframe the present points to fit the preferred model.

A final and especially important locutional note: I employ ‘experience’ in a somewhat non-standard way, as an attitude-neutral term that can pick out any particular bit of mental life. The term is attitude-neutral in the sense that it can pick out beliefs, desires, hopes, emotions, perceptual states, or any other attitude kind. Another note about ‘experience’ is that it does not necessarily denote one kind of attitude but can refer to bits of mental life that are comprised by multiple distinct attitudes. Thus, we can talk of the experience that is comprised of *seeing something, judging it to be a piece of chocolate cake, and desiring to eat it*. The reason we require this flexible locution is that the main question of this essay is whether certain relevant experiences might be explained by certain visual-cognitive combinations or whether they might be explained by a visual state alone. Hence, we require a term that is neutral on this question.

³² Burge proposes this particular non-propositional structure in Burge (2010).

1 What Kinds of Features Can We See?

Recently, the range of proposed high-level percepts has expanded dramatically. Some of these include: artifactual features (e.g., *being a guitar*), first-personal agency, dispositional features, absence, event-causal features, natural kind features (e.g., *being a lemur*), a range of social features (*being a face*, *being masculine*, *being an expression of happiness*), and moral features (Audi, 2013; Bayne, 2009; Bayne, 2011; Block, 2014; Butterfill, 2009; Chappell, 2008; Cullison, 2010; Farennikova, 2013; Fish, 2012; Nanay, 2011; Scholl & Gao, 2013; Scholl & Tremoulet, 2000; Siegel, 2005; Siegel, 2009; Siegel, 2010; Sorenson, 2012; van Gulick, 1994; cf. Byrne, 2009, Chalmers, 2006; Logue, 2012; Pautz, 2009; Prinz, 2013).

Strictly speaking, whether some feature is high-level or low-level is always relative to a sensory modality. For the visual modality, high-level features are any features other than the low-level features, whose paradigmatic instances are shape, color, edge, contrast, form, and movement features. For the auditory modality, high-level features are any features other than the low-level features, whose paradigmatic instances are pitch and loudness; and so on for the other modalities.

Thus, the claim that we can visually perceive a thickly frosted piece of cake as sweet counts as a high-level claim, since *being sweet* is not among the visual modality's low-level features. But the claim that we can taste a piece of cake as sweet does not count as a high-level claim since *being sweet* is (arguably) among the gustatory modality's low-level features. Though the fact of whether some feature is high-level is, strictly speaking, always relative to a sensory modality, the majority of high-level percepts commonly posited count as high-level relative to any modality. For instance, *being a train* and *being the cause of a certain effects* are high-level no matter what modality they are represented in.³³

The recent trend in philosophy of perception of expanding the range of features that can serve as percepts will (and on my view, should) have dramatic implications for philosophy of psychology in general. To take just one potential area of impact, philosophy of psychopathology stands to gain from ‘more’ perception a greater number of plausible hypotheses about various psychological disorders. For instance, it has been suggested that *thought insertion*, one of the hallmarks of schizophrenia, is at least partly constituted by an abnormal perception of thought agency;³⁴ hence, the possibility that a high-level feature such as the feature of *being the cause of one's thought* is represented in perception has already resulted in a novel hypothesis about long-established empirical data.

³³ Notably, the low-/high-level distinction cross-cuts the distinction between early and late stages of perceptual processing. In principle, low-level features might be processed at a late stage and high-level features might be processed at an early stage.

³⁴ Here's one description of a patient report that suggests that a perceptual account has some advantages over its cognitive competitors:

[S]he said that sometimes it seemed to be her own thought “. . . but I don't get the feeling that it is.” She said her “own thoughts might say the same thing . . . but the feeling isn't the same . . . the feeling is that it is somebody else's . . .” She was asked if she had other people's thoughts put inside her head. She said “. . . possibly they are but I don't think of them in that way . . . they were being put into me into my mind . . . very similar to what I would be like normally” (from an unpublished 1999 manuscript by Allison-Bolger, excerpted in Horerl, 2001).

Of particular interest is that this patient does not disown the *content* of the thoughts she describes as feeling like ‘somebody else's,’ since these thoughts are ‘very similar to what [she] would be like normally’ and ‘her own thoughts might say the same thing.’ Apparently, then, thought insertion is not simply a matter of patients observing themselves to have thoughts that are inconsistent with their moral or epistemic commitments and then *inferring* or *judging* that those thoughts are not theirs. As a result, any view which attempts to explain the phenomenon in terms of a cognitive response to the foreignness of some *content* does not seem very promising.

On a different front, if perceptual space were generally enlarged to encompass a very diverse number of high-level features, this may have implications for the kinds of explanations required by accounts of justification of external world beliefs. Consider for instance foundationalism about epistemic justification, according to which beliefs about the external world must be justified either directly, by perceptual states, or indirectly, by other beliefs that are themselves perceptually justified. For the foundationalist, it very much matters which features are represented in perception. For on a conservative drawing of the perceptual landscape—on which, say, only shape and color are represented in perception—a foundationalist must explain how our very many beliefs about the external world derive justification from a relatively impoverished base of perceptual experiences (or else give up some of these beliefs to the skeptic). Conversely, on a more generous drawing of the perceptual landscape—one which might include things such as the feature of being an elm tree, the feature of being angry, or the feature of being the cause of another event—the foundationalist's workload looks to be somewhat lighter. This isn't to say that high-level perceptual disputes directly bear on the plausibility of foundationalism. But their outcome does directly influence what the foundationalist must explain (Macpherson, 2011; Masrour, 2011).

Not only might a generally enlarged perceptual space have implications for the foundationalist account of external world knowledge, a number of particular high-level disputes may have implications for particular philosophical disputes. To give just one example: the dispute over whether perception can represent *singular* features (e.g., being *that* chair) has been thought by some metaphysicians of perception to play an important role in the debate over disjunctivism. According to disjunctivism, hallucinatory mental states—even those indistinguishable from veridical perception—are not of the most fundamental kind to which perceptual states belong. Some disjunctivists have attempted to shore up support for their view by arguing first, that perception

represents singularity and second, that disjunctivism is better suited than its competitors to account for this fact (Tye, 2009). Clearly, if this strategy is to work, some evidence is required that singularity *is* perceptually—and not merely cognitively—represented.

Philosophical interest in perception has tended to focus on perception as a source of epistemic justification and also, as a source of evidence for various metaphysical positions. Consequently, philosophical interest in high-level perceptual disputes has tended to focus on the purely existential questions of whether high-level features are (ever) represented in perception and, if this question is answered in the affirmative, which high-level features are (ever) represented in perception. Further questions about high-level perception—such as the question of its ubiquity and relatedly, of the circumstances that elicit it—philosophers have tended to leave in the hands of psychologists.

On my view, the question of the ubiquity of high-level perception is at least as philosophically interesting as the purely existential questions. Consider, by way of example, the perception of agency. Supposing that agency is sometimes represented in perception, we might still wonder how common perception of agency is in ordinary experience. Is it, as some have suggested, a nearly ever-present feature of experience (Siegel, 2005; Strawson, 2011)? Or is it exceedingly rare, present perhaps when one loses one's virginity or leaves the family business, but not when one pours a cup of coffee or opens a door? If the importance of philosophy of perception lies not only in its special relationships with epistemology and metaphysics, but also, and perhaps foremost, in its ability to shed light on human experience and behavior, it should matter to philosophers not only whether, but also when, high-level features are perceptually represented.

Of course, it is not only ordinary human experience and behavior that philosophy of perception promises to illuminate, but also highly atypical, even pathological, experience and behav-

ior. Returning again to the example of perception of agency, it seems plausible that agency perception—or its lack—might figure in a number of disorders, such as substance abuse, obsessive-compulsive disorder, and bipolar disorder. Consider in particular the bipolar patient in the grip of a manic episode. Could it be that this person suffers from an *absence* of agency perception that accounts for the impulsivity that is characteristic of mania? Or might she rather suffer from an *excess* of agency perception, enjoying agency perception under conditions that would not elicit it in normal subjects, and that this accounts for the delusions of grandeur that are also characteristic of mania? (Moeller et al., 2001) Or is it some combination—is her mental state in some respects abnormally devoid of agency and in other respects abnormally awash with agency perception? If the answers to these questions are important, then it matters not merely whether but also when high-level features are perceptually represented.

2 Introspective Strategies

In this section, I describe two broadly introspective arguments in favor of high-level perception, both from the recent literature. These are the arguments from direct introspection and from veridicality judgments.

2.1 Introspection and High-level Perception

Thus far, introspection of disputed cases has not proven to be helpful in settling disputes about high-level perception. Bayne aptly describes the kinds of stand-offs that are common:

... liberals claim that introspection reveals clear instances of high-level perceptual phenomenality, whereas conservatives deny that this is so. Look at a tomato, the liberal says, is there not something it is like to see it as a tomato? The conservative shakes his head in puzzlement (Bayne, 2009, p. 390).

Nevertheless, advocates of high-level perception—including Bayne—have occasionally appealed to introspection in a qualified and partial way, as a source of weak, defeasible evidence in

favor of various kinds of high-level perception. For instance, in the context of arguing in favor of the perception of one's own agency, Bayne describes a vignette in which one runs out of a burning building, making several decisions along the way about how best to escape. Bayne thinks that reflecting on the subject's experience in this case suggests at least a prima facie reason to endorse the perception of one's own agency:

The central challenge facing the sceptic is to explain (away) the appeal of such vignettes as [the fire vignette]. The mental state of oneself as an agent is, for many of us, a robust and readily identifiable phenomenon. The sceptic needs to explain why it is so plausible to *think* that there is a phenomenology of agency even though there isn't (Bayne, 2008, p. 185).

Similarly, Dustin Stokes suggests introspection provides prima facie evidence in favor of the claim that we can visually perceive aesthetic features, such as *being somber* and *being vivid* (Stokes, 2014, p. 2293). Anna Farennikova (2014) suggests introspection provides prima facie evidence in favor of the view that we can visually perceive absences. Though introspective evidence is in all of these instances taken to be weak and defeasible, I will ultimately argue that even this extremely modest reliance on introspection is misguided. This is for the reason that while introspection can help us determine the content of a disputed experience, introspection is silent on the further issue of whether the vehicle of that content is itself perceptual or is rather a post-perceptual judgment or thought.

2.2 *The Veridicality Method: Tye's Argument for Particularism*

A second strategy that has been used to argue for various high-level perceptual claims relies on veridicality judgments. On this strategy, we first start with a target experience and think about the range of cases in which this experience would be accurate and those in which it would be inaccurate. If some experience would only be accurate in cases in which some particular high-level feature Φ obtains, this suggests a reason for thinking that Φ is perceived.

For instance, consider the typical subject who views a spoon, in normal viewing conditions. Plausibly, this subject's experience is accurate only if there is a spoon in front of her. It won't be accurate if there is merely a spoon-looking light display before her. Arguably, this suggests some reason for thinking this subject enjoys a visual experience as of a spoon. Michael Tye employs a veridicality argument in precisely this manner, to defend the view that the objects of perception are singular and not merely existential.³⁵ On the 'singular' view of perception, the objects of perception are particulars, such as *that red blob*. On the 'existential' view of perception, the objects of perception are merely existentially quantified; perception can take *a red blob* as an object, but it cannot take any *particular* red blob as an object. In favor of the view that particulars are perceived, Tye considers the following case:

Suppose that I am looking directly ahead and that, unknown to me, there is a mirror in front of me placed at a 45 angle, and behind which there is a yellow cube. To the right of the mirror, and reflected in it, is a white cube. Through special lighting conditions, this cube appears yellow to me (Tye, 2009, p. 79).

Relying on this case, Tye argues as follows: the subject's experience gets the world wrong in some sense. In particular, it misrepresents the color of the particular cube that is the object of his experience—this is the cube that is to the right of the mirror and projected onto the mirror. *This* cube is really white, but the lighting makes it look yellow. If experience represents particular objects, then we can accommodate the result that this experience is non-veridical; this experience is about *that* cube, the one to the right, and it is in misrepresenting this cube's color that the experience is inaccurate. In contrast, the view that experience can only represent existential objects cannot account for the fact that this experience is intuitively inaccurate. On the existential view,

³⁵ Tye does not endorse a representational view of perception's objects, so my locution here reflects that fact.

one's visual experience merely represents *a* yellow cube, and there is a yellow cube before the subject, namely the one occluded by the mirror (Tye, 2009, pp. 79-80).

My position on the veridicality method is essentially the same as my position on the purely introspective method. The veridicality method is useful in settling high-level disputes, but only at the level of *content*. It does not help settle the further question of whether it is perception or merely post-perceptual judgment that is the *vehicle* of that content. The reason Tye's argument is attractive is that it does provide evidence for something, namely that the described experience represents singularity *somehow or other*. But contra Tye, the argument provides no evidence whatsoever that singular objects are the objects of *perception*, since it does not exclude the possibility that singularity is represented merely at the level of a post-perceptual judgment.

3 A Limitation on Introspection: Attitudinal Opacity

In this section, I argue that introspection provides no evidence in favor of high-level perception. The reason for this is that the kinds of experiences which are thought to involve high-level perception are *attitudinally opaque*: while introspection on these experiences can provide evidence of the content of those experiences, introspection is silent on the further question of which attitude is the vehicle of that content. In other words, introspection can tell us that (say) *being a face* is represented somehow in experience, but introspection is neutral on the further question of whether *being a face* is represented by vision or merely at the level of a post-perceptual judgment.

3.1 Attitudinal Opacity

The claim that introspection is, in the relevant cases, opaque about which attitudes are involved is not the claim that introspection is in these cases misleading or unreliable. It is not, for instance, the claim that in some cases, introspection 'says' that one's mental state is perceptual when it is not, or that it 'says' one's mental state is not perceptual, when it is. Rather, attitudinal

opacity is the view that in some range of cases, introspection is simply silent as to whether the relevant state is perceptual.

Relatedly, the claim that in the relevant cases, introspection is attitudinally opaque is not the claim that introspection can tell us *nothing* about the attitude(s) which comprise the relevant experience. It may be that introspection can tell us that the relevant experience is not intense joy. Thus, introspection may rule out certain attitudinal claims. Moreover, introspection probably can tell us that the relevant experience involves visual perception of features like shape and color. That is to say, introspection probably does provide evidence for low-level perception. But introspection seems to be silent about whether high-level contents, such as being a face, or represented perceptually or merely cognitively.

My argument that the disputed experiences are attitudinally opaque proceeds in three steps: first, I draw on empirical evidence to argue that some judgments are fast, automatic, and, though themselves conscious, are the result of non-conscious processes; second, I suggest that the most viable cognitive alternative to the claim that there is high-level perception is one which appeals to these fast, automatic, non-consciously processed judgments. Third, I argue that once we understand the relevant cognitive explanation along these lines, reflection on the relevant cases reveals these cases to be attitudinally opaque.

Some judgments are slow, effortful, and the result of conscious deliberation. For instance, if you calculate how to split a complicated bill between many parties, your resulting judgment about how to split the bill is probably processed relatively slowly, accompanied by felt effort, and mediated by inferences which are themselves conscious. However, many judgments are formed quickly, without effort, and—though themselves conscious—are the result of non-conscious processes (Evans, 2008). For instance, consider the following math problem; answer it as quickly as you can:

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? Answer: ____ minutes

If you are like most people, the answer ‘100’ probably popped into your head (Thompson, 2009). However, the correct answer to the question is *5 minutes*.³⁶ That most participants give an incorrect answer is somewhat surprising, given that the correct answer can be arrived at using a bit of simple algebra.

The standard explanation for the fact that the incorrect answer immediately comes to mind is that this answer is arrived at by something like a *matching heuristic*: the system notices the pattern 5:5:5 and the partially completed pattern 100:100: ____ and generates ‘100’ as the correct answer.²⁹

I say that ‘the system’ notices the relevant pattern and not ‘the subject’ because it’s highly unlikely that the typical subject applies the matching heuristic by a process of explicit inference, or by constructing an argument. This is for two reasons. First, inference is time-consuming, and the incorrect answer is produced very quickly. Second, had the subject engaged in inference, it’s unlikely she would have made the mistake of applying the matching heuristic to the ‘widgets’ problem, since a moment’s consideration shows the application of this heuristic to be inappropriate to the context. Rather, on the best explanation, the attractive, incorrect judgment is the conscious output of a process which is itself non-conscious. Relatedly, the incorrect answer to the problem is not associated with effortful deliberation or a feeling of initiating some process. One simply ‘sees’ the answer. The ‘widgets’ problem is just one of many cases that tend to elicit automatic, fast, and non-consciously processed judgments (Evans, 2008).

³⁶ The trick is to think of the rate at which the machines generate widgets. 5 machines work at a rate of 1 widget per minute. Therefore, 100 machines work at a rate of 20 widgets per minute, which means it will take 100 machines *5 minutes* to complete 100 widgets.

The most attractive version of the cognitive alternative to high-level perception will appeal to judgments that are formed quickly, automatically and as the result of non-conscious processes. This version of the view can accommodate the fact that, when viewing a person across from you, it just *seems* that there is a face before you.

On a cognitive view which appealed to slow, effortful, consciously processed judgments, it would be very difficult to explain the fact that, in the 'face' experience, it just *seems* to you that there is a face. This is because we don't seem to experience any temporal gap between the visual perception of shape and color and the experience as of a face. We also don't experience anything like effort in generating the 'face' experience; nor are we aware of any inferential steps mediating the perception of shape and color and the 'face' experience.

The version of the cognitive view which appeals to fast, automatic, non-consciously processed judgments can explain all of the preceding subjective aspects of the 'face' experience. The reason there is no discernible temporal gap between the visual perception of shape and color and the onset of the 'face' judgment is that the relevant judgment is generated very quickly after the onset of the visual perception of shape and color. The reason there is no feeling of effort or initiation accompanying the 'face' judgment is that this judgment is formed without conscious initiation or effort. Finally, the reason there is no conscious access to the kinds of inferences or processes which mediated the judgment is that these processes are non-conscious, though their output, which is the judgment itself, is conscious. In sum, by appealing to a fast, automatic, non-consciously processed judgment, the cognitive explanation is well-positioned to account for the fact that for the subject viewing a face, it just *seems* that there is a face.

That there is a version of the cognitive explanation which can account for some of the subjective aspects of the 'face' experience is not itself sufficient to show that introspection does not adjudicate between the cognitive and perceptual explanations of the experience. For one thing,

there may be further subjective features—beyond those of temporal onset, automaticity, and mediating processes—which do tell against a cognitive explanation.

It may be that introspection tells against the cognitive explanation via some sort of mechanism that doesn't rely on phenomenal features. For instance, Georges Rey has recently proposed that experiences are 'tagged' by the kind of sub-routine, module, sensory modality, or other process that produced them and that these 'tags' form the basis of our knowledge of our occurrent attitudes (Rey, 2013). Moreover, these 'tags' do not necessarily correspond to a phenomenal feel. Rey argues from the case of *efferent copies*, which appear to contribute to knowledge of one's own physical movements. Rey argues that efferent copies are 'tagged' as having been generated by the motor system and in this way afford knowledge such as that one is moving one's eyes to the right. As Rey suggests, 'these efferent copies don't seem to give rise to any phenomenal experiences. One 'just knows', e.g. that one is moving one's eyes on the basis of these commands alone' (Rey, 2013, p. 279). If Rey is correct that self-knowledge of one's attitudes is due to 'tags' which themselves lack a phenomenal feel, this suggests that introspection might adjudicate between perceptual and cognitive explanations of the 'face' experience, in a way that does not exploit phenomenal differences between those explanations.

I will argue that, as it turns out, introspection does not adjudicate between the perceptual and cognitive explanations of the 'face' experience, whether through phenomenal feature or through Rey-style non-phenomenal 'tags.' This is a contingent fact about human psychology; had we been different, we might have enjoyed introspective knowledge of the relevant attitude facts, just as had we been different, we might have been such that we could introspectively access facts about the neural processes that realize our occurrent states. But just as we cannot determine, via introspection alone, which neural processes realize our occurrent states, we cannot tell whether high-level experiences are perceptual or are merely post-perceptual judgments.

I should advertise in advance that the argument I propose in favor of attitudinal opacity itself relies on introspection. However, in the present context, that I argue for a certain introspective limitation by appealing to introspection is not problematic. This is because attitudinal opacity is not the claim that introspection leads to *error*; it's rather the claim that introspection is sometimes silent about certain attitudinal facts. Thus, even if true, the claim that introspection is attitudinally opaque does not suggest a reason to doubt the verdicts of introspection.

Consider again the experience of viewing a person across from you, under normal conditions. Now here is the question: can you determine, just from reflecting on that experience, whether it involves visual perception as of *being a face* or merely a fast, automatic, non-consciously processed post-perceptual thought, 'that's a face', a judgment which is elicited in response to low-level visual information? It seems to me that if such a question were put to me, I would not be able to answer it on the basis of introspection alone. Both suggested explanations of the experience seem to me to be *compatible* with the verdicts of introspection.

Further, it seems to me that introspection *does* suggest that my experience is in some sense about a face. I could not mistake this experience for one that is about a lamp or the Antarctic. This introspective evidence that my experience is about a face may be weak or defeasible. Thus, in this test case of face experience, introspection seems to be at least somewhat useful in helping settle that the experience is about a face but silent on the further issue of whether this 'face' content is represented by perception or by a post-perceptual thought.

It also seems to me that introspection does provide evidence that at least some aspects of the face experience are visual. In viewing the face, I could not seriously wonder if my experience is olfactory or tactile; I know it involves visual perception. But the further issue of whether the 'face-ness' of the experience is visually represented is something introspection simply does not address.

3.2 Criticizing Introspective Strategies

If the relevant kind of attitudinal opacity holds for the very experiences which are meant to motivate high-level perception, then introspective considerations cannot offer even weak, defensible evidence in favor of high-level perception. Further, this reliance on introspective evidence has a less-than-salutary effect on the entire debate. It bestows on the perceptual hypothesis a prima facie plausibility the view does not deserve. Viewed in the right light, the debate between the perceptual and cognitive explanations of experiences like the ‘face’ experience is a debate about how to individuate two very important attitudes—perception and cognition—in a theoretically satisfying way. It is emphatically not a dispute about how rich phenomenology is or about what some experience is ‘like.’ Both sides can agree that phenomenology is rich *to exactly the same degree*; what they disagree about is which complex of mental attitudes explains this phenomenology.

If perception is attitudinally opaque, attempts to argue for high-level perception by appealing to veridicality judgments also do not succeed. This is because this strategy indirectly relies on introspection. Recall that the veridicality method takes at face value judgments about when a mental state would be (non-)veridical. So for example, in Tye’s case of the experience of the illusorily yellow cube, we would intuitively judge this experience to be inaccurate in a certain respect, because it misrepresents the color of the particular cube that is the object of that experience. Plausibly, this experience is non-veridical even though there is a yellow cube in the vicinity, namely the one occluded by the mirror. Tye argues that this (non-)veridicality judgment suggests that the object of this experience is the particular cube projected onto the mirror, and hence that perception’s objects are particulars.

Tye’s argument does establish that particulars are somehow represented in experience. The problem is that it does not establish that particulars are objects of visual perception. The

argument provides no reason whatsoever that might exclude a cognitive explanation of particularity, on which perception's objects are merely existentially quantified, and particularity is supplied at the level of a post-perceptual thought, e.g. '*that* cube is yellow.' It is consistent with the experience's being somehow inaccurate that its inaccurate component is not perception itself but rather a relevant post-perceptual thought.

In effect, Tye takes for granted that it is in virtue of visual perception that particularity figures in the experience. He never considers the alternative view on which visual perception's objects are merely existentially quantified, and particularity is introduced at the level of post-perceptual judgment. It seems reasonable to think that he has supposed the attitude question to be trivially or easily settle-able, presumably by introspection. If the considerations I have presented in favor of attitudinal opacity are correct, this would be a mistake.

This concludes my criticism of broadly introspective strategies of arguing for high-level perception. In the next two sections, I consider and reject two other approaches to arguing in favor of high-level perception. I first consider an argument, due to Susanna Siegel, that certain experiences exhibit a *felt unity* that would be difficult to explain without positing that these experiences involve certain kinds of high-level perception. I second consider an argument, due to Tim Bayne, that an adequate explanation of associative form visual agnosia requires that we posit that typical, non-agnosic subjects sometimes visually perceive artifactual features, such as *being a stethoscope*.

4 What Felt Unity Doesn't Show: Reply to Siegel

In a series of papers and in a monograph, Susanna Siegel has argued that we visually perceive a range of high-level features, including natural kinds (e.g., *being a lemur*), event-causal features, affordances (e.g., *being to-be-eaten*), and our own agency (Siegel, 2005; Siegel, 2009; Siegel, 2010; Siegel, 2014). Siegel's argumentative strategy is broadly consistent across all of these

cases: for some relevant high-level feature Φ she considers a carefully chosen pair of phenomenally contrastive experiences and argues that the phenomenal difference in these experiences is due to a difference in content, such that one of these experiences represents Φ , whereas the other does not. Siegel then argues from inference to the best explanation that the experience as of Φ is *perception* as of Φ , as opposed to some other attitude, such as a judgment, thought, or assumption. In this way, Siegel's argumentative strategy appropriately distinguishes between the content claim that experience represents Φ somehow or other, and the attitudinal claim that Φ is represented at the level of perception, instead of by some other kind of attitude.

One argument that Siegel employs to argue against a cognitive explanation of Φ -experience is that cognitive explanations cannot accommodate a kind of *felt unity* between low-level visual information—such as information about shape, color, and movement—and the associated high-level feature. In this section, I argue against Siegel's 'felt unity' objection to cognitive explanations, on the grounds that low-level visual experiences combined with fast, automatic, effortless, and apparently unmediated thoughts can elicit a feeling of unity, particularly if the low-level visual experience and the associated thought are *narratively coherent*. Thus, the fact that the experiences Siegel considers are marked by a felt unity does not suggest a reason to exclude a cognitive explanation of such experiences.

4.1 Siegel on Felt Unity

To illustrate Siegel's strategy of arguing for high-level perception from an experience's felt unity, we can focus on her argument in favor of visual perception as of event-causal features (Siegel, 2009; Siegel, 2010). Siegel argues that we see event-causal features in a range of ordinary cases. For instance, when we see a cat lounging in a hammock, we visually perceive the pressing down of the cat's body as causing the downward pull of the hammock (Siegel, 2009). Siegel's discussion of the 'felt unity' argument focuses on the following experience of causation:

Suppose that you are playing catch indoors. A throw falls short and the ball lands with a thump in a potted plant, with its momentum absorbed all at once by the soil. You see it land, and just after that the lights go out. The ball's landing in the plant does not cause the lights to go out, and let us suppose that you don't believe that it does. Nonetheless, it may seem to you that the ball's landing somehow caused the lights to go out (Siegel, 2010, p. 122).

On Siegel's preferred explanation of the described case, the observer visually perceives the ball's landing as causing the lights to go out. This view contrasts with a view on which the observer merely visually perceives low-level features such as shape, movement, color, and perhaps objecthood, and on the basis of these features quickly generates a post-perceptual thought, 'the ball's landing causes the lights to go out.' On this alternative, visual-cognitive view, the post-perceptual thought is in conflict with the observer's better judgment that there is no causal connection between the ball's landing and causing the lights to go out.

One argument Siegel offers against the cognitive explanation is that there is a felt unity in the described experience that the cognitive explanation cannot accommodate. In particular, the connection between the visual experience as of low-level features and the causal interpretation are deeply intertwined, fluid, and appear to be one and the same experience. Siegel argues that so long as the visual experience and the subsequent thought are not constitutively connected, and so are genuinely distinct attitude tokens, the cognitive explanation cannot do justice to the unity of the 'ball' experience:

One option is that the sensory component causes the causal representation, without there being any law-like connection between the property of being that kind of sensory state and the property of being a belief or disposition to believe with the specific causal content posited . . . But such a causal link does not seem to be able to account for *how closely the cognitive component would be felt to be associated with the sensory component* (Siegel, 2009, p. 534, emphasis mine).

The kind of felt unity Siegel claims is present in the 'ball' case is different from the mere unity of

subject consciousness that has sometimes been argued to be an aspect of all experiences had by the same subject (see for example, Bayne, 2010). The kind of unity Siegel is after can be illustrated ostensively: whereas the ‘ball’ experience enjoys this unity, it is not had by a pang of sadness that somehow causes one to remember to buy light bulbs (Siegel, 2009, pp. 534-535).

4.2 Felt Unity from Narrative Coherence

I take Siegel to be correct that there is a kind of felt unity in the ‘ball’ experience, of the kind that is plausibly absent when a pang of sadness causes one to remember to buy light bulbs. But I would suggest that the visual-cognitive explanation of the ‘ball’ experience—and of event-causal experiences more generally—enjoys the resources to account for this felt unity. In particular, in at least some cases where two different mental states are both: attended, near-simultaneous in onset, temporally overlapping, and narratively coherent, a feeling of unity can arise. It is not required that these mental states be of the same attitude kind. Due to the *ceteris paribus* nature of psychological laws, these conditions may not universally suffice for a feeling of unity. But the fact that these conditions can sometimes elicit a feeling of unity shows that the visual-cognitive explanation of the ‘ball’ experience might also lead to a feeling of unity, even though it posits that the experience is due to two different mental attitudes, one a visual state and one a thought.

In the present sense, *narrative coherence* obtains between two or more mental states when those states are semantically linked with respect to their contents. One way two contents can be narratively coherent is to be at least partly about the very same state of affairs, where this state of affairs is represented under the same mode of presentation (if any mode of presentation is in play). For instance, a state that represents a particular cat as lounging in a hammock and a state that represents the hammock as in need of repair are narratively coherent in virtue of both taking as an object *the hammock*. There may be a psychologically important sense of narrative coherence on which states that are not even partly about the same states of affairs can count as

narratively coherent, but for our purposes, we can focus on narrative coherence qua shared intentional objects.

It turns out that mental states that are both attended, apparently unmediated by any other mental state, of simultaneous or near-simultaneous onset, and both at least partly about the same state of affairs can elicit an overall feeling of unity. In order not to beg questions about whether high-level experiences are perceptual, I will focus on cases that do not involve visual perception. Rather, I focus on cognitive-emotional experiences and show that in at least some cases, they elicit a feeling of unity.

First, suppose you learn that your sister is pregnant with her first child. You may feel joy for her, mixed with excitement at the prospect of having a new nephew or niece. It seems plausible that your judgment that your sister is to have a child might be felt to be unified with your joy and excitement, and at least part of what explains this is their shared intentional object of your sister's pregnancy. Second, suppose you should learn that your roommate's pet tarantula has escaped his terrarium and is wandering somewhere in your apartment. Phobic of spiders, you are immediately filled with terror. Plausibly, the judgment that there is a tarantula roaming unhampered in your home and the terror are felt to be unified, and at least part of what explains their felt unity is their shared object of your roommate's tarantula.

If the judgment that your sister is pregnant can be felt to be unified with subsequent joy, and the judgment that your roommate's tarantula has escaped his cage can be felt to be unified with subsequent terror, it seems plausible that in the 'ball' case, your visual experience as of the ball's trajectory and the subsequent dimming of the houselights might be felt to be unified with a subsequent thought that the landing of the ball causes the dimming of the houselights. At least, that your visual experience and your judgment of causation are of distinct attitude types is no barrier to such a feeling of unity.

On the visual-cognitive explanation of the felt unity in the ‘ball’ case, the felt unity of the visual experience as of shape, motion, and objects is felt to be unified with the thought of causation for the reason that the visual experience and the thought it elicits are near-simultaneous in onset, partly temporally overlapping, and partly share intentional objects. If the relevant thought is of the form ‘that causes that’, where each ‘that’ demonstratively refers to a different seen event (the ball’s landing or the lights’ dimming), both the visual state and the post-perceptual thought are at least partly about certain events (it’s less clear whether these events are represented under an ‘events’ presentation). I conclude that the felt unity in the ‘ball’ case does not suggest a reason to rule out a partly visual-cognitive explanation of that experience.

In Siegel’s most recent defense of the view that we see causal features, she offers a more direct argument against the cognitive explanation of the event-causal experience, viz. that even if the visual experience as of the ball landing in the plant and the houselights going out were relevantly associated with a thought of causation, this would not itself suffice to reinsert the feeling of unity between these visual aspects of the experience and the causation interpretation:

Throughout the day, you might be periodically struck by the thought that a certain ball’s landing in a plant with a thud caused the lights to go out . . . And on one of those occasions, you’re struck with the thought just as the lights really do go out, just after the ball you’ve been throwing (and musing about) lands with a thud. All of this could happen without it seeming to you that the ball’s landing in the plant was unified in the way we’ve been discussing with the light’s going out (Siegel, 2010, p. 138).

If combining the relevant low-level visual experience with a relevant demonstrative thought about causation could not recreate the feeling of causation present in the original case, I agree that this would be a reason to reject the cognitive explanation of event-causation. However, the proposed test condition is a very bizarre one: it would have to be one in which we happen to demonstratively ostend two events (the ball’s landing and the disillumination, respectively) and think of them as bearing a causal connection at nearly the same moment that we visually perceive

the two events as obtaining. There is very little in typical experience that might permit us to judge with any confidence whether such a visual-thought configuration might result in a seeming of causation.

Further, given the felt unity in the pregnancy case and the tarantula case, we have some precedent for thinking that a low-level visual experience combined with a relevant causation-thought *would* elicit a feeling of unity. The felt unity in the pregnancy and tarantula cases is plausibly explained by the fact that these cases each involve states of near-simultaneous onset, partial temporal overlap, and partially shared intentional objects. Since the test case of low-level visual experience combined with a relevant causation-thought would also meet these criteria, it would not be surprising if it too should exhibit the unity of the original case.

5 What Agnosia Doesn't Show: Reply to Bayne

In two papers, Tim Bayne has argued that the best explanation of visual form associative agnosia requires that we posit high-level perception of artifacts (Bayne, 2009; Bayne, manuscript). On this view, associative agnosia involves intact perception of low-level features such as form and movement and an abnormal absence of perception of artifactual features such as *being a stethoscope* or *being a pipe*. The form of agnosia Bayne considers is characterized by intact visual perception of shape, color, and movement but deficits in the recognition of visually-presented objects. Here is a case report of one such patient:

When shown a stethoscope, he described it as 'a long cord with a round thing at the end', and asked if it could be a watch. He identified a can opener as 'could it be a key?' Asked to name a cigarette lighter, he said, 'I don't know', but named it after the examiner lit it. He said he was 'not sure' when shown a toothbrush. Asked to identify a comb, he said, 'I don't know'. When shown a large matchbook, he said, 'It could be a container for keys' (Rubens & Benson, 1971).

Pure visual form associative agnosia (henceforth: ‘agnosia’) does not appear to involve a general cognitive deficit, as evidenced by the fact that agnosics can successfully identify objects presented to them via non-visual modalities. For instance, the same patient described in the passage had no difficulty identifying objects presented to him through touch (Rubens & Benson, 1971). Bayne’s argument is a kind of phenomenal contrast argument: deficits in low-level visual perception cannot explain the agnosic’s deficit, as his visual perception of low-level features is intact. What distinguishes the agnosic from his typical counterpart is that the agnosic lacks the ability to visually perceive objects as stethoscopes, combs, can openers, lighters, and matchbooks, whereas his counterpart visually perceives objects as stethoscopes, combs, can openers, lighters, and matchbooks.

5.1 Bayne’s Argument for Perception of Artifactual Features

The relevant portion of Bayne’s argument might be stated as follows:

- (1) The agnosic’s deficit in object recognition is not due to low-level visual impairments.
- (2) On the best explanation, the agnosic’s deficit in object recognition is due to an absence of artifactual visual perception.

In arguing for the first premise, Bayne follows much of the literature on associative agnosia, which treats the associative form of agnosia as characteristically involving intact low-level perception. Here is Bayne on this point:

. . . what kind of perceptual content has the patient lost? He has not lost low-level perceptual content, for those abilities that require the processing of only low-level content remain intact. The patient’s deficit is not one of *form* perception but of *category* perception. Hence high-level perceptual representation – the representation of an object as a stethoscope, a can-opener or a comb – can enter into the contents of perceptual phenomenality (Bayne, 2009, p. 391, emphases his).

The problem with Bayne's argument is one of (inductive) validity: that the agnosic's deficit is not to do with processing of low-level visual features does not suggest his deficit to be best explained as a deficit in the perceptual processing of artifactual information. There exists an important alternative explanation of the agnosic's deficit, on which his characteristic deficit is due to a deficit in the *simultaneous visual processing of multiple spatial relations*. On this view, the core deficit of agnosia is a deficit in holistic processing. An impairment at this level promises to explain both why agnosics exhibit typical low-level visual capacities and why they cannot categorize visually-presented objects: local visual experience is normal in the agnosic, but his ability to visually integrate multiple spatial features into a composite is compromised, which leads to deficits in object categorization. Further, as I will now argue, there is good evidence to suggest that associative agnosics suffer from a holistic processing deficit.

5.2 Associative Agnosia As a Deficit of Holistic Processing

The literature on agnosia distinguishes between apperceptive agnosia and associative agnosia: apperceptive forms characteristically involve deficits in low-level visual processing whereas associative forms characteristically do not involve low-level deficits visual processing. However, this characterization is silent on the issue of whether associative agnosia might be due to a deficit in intermediate processes, such as in the ability to simultaneously integrate multiple local features into a coherent visual whole. Farah (1990) reviewed 99 cases of associative agnosia and argued that the few cases that were sufficiently tested for such holistic visual deficits revealed them, and that all other cases lacked adequate tests of such deficits. Farah hypothesized that such deficits might account for the categorization deficits that are characteristic of these forms of agnosia.

After Farah's 1990 work, several subsequent cases of associative agnosia were described in the literature, but none tested Farah's hypothesis that associative agnosia necessarily involves

holistic perceptual deficits that explain the object recognition deficit (Delvenne et al., p. 598). It was not until 2004 that a team led by Jean-François Delvenne tested Farah's hypothesis by carefully investigating holistic capacities in a subject with associative form agnosia (Delvenne et al., 2004). This subject exhibited very good visual acuity, a full visual field, normal vision as measured by several traditional tests, and a preserved ability to copy presented figures (Delvenne et al., 2004, pp. 599-600). Nevertheless, when tested on two tasks designed to measure holistic processing, this subject performed worse than controls, suggesting an impaired ability to visually integrate an object's structural features.

On the first task, the patient was presented with a series of depicted objects and asked to determine whether the object was 'impossible' or 'possible.' Some presented objects were 'impossible' in the sense that their local spatial features could not simultaneously obtain in the same object (These objects are comparable to the artist Escher's impossible staircases, though unlike Escher's staircases, these objects were not familiar objects). Some presented objects were 'possible,' in that their local spatial features could simultaneously obtain in the same object. The agnosic patient performed worse at this task than typical controls, a fact the authors took to indicate a deficit in 'the complex integration of object parts at the level of the three-dimensional structural description system' (Delvenne et al., 2004, p. 604-605).

On the second task, the agnosic was presented with pairs of depicted novel objects which differed only in subtly different configurational features. For instance, one pair of objects were identical except that one object's 'arms' pointed upwards, whereas the other object's 'arms' pointed downwards. Instructed to indicate whether a presented pair was spatially identical or non-identical, the agnosic again performed worse than typical controls (Delvenne et al., 2004, pp. 605-06). Given the subtlety of the configurational differences between the presented objects, the task would benefit from an ability to simultaneously process multiple spatial features. Thus, a

plausible explanation of the patient's performance on this task is a deficit in the capacity to holistically visually process objects.

It is important to appreciate that the deficit in holistic processing supported by the agnosic's performance on these tasks cannot be explained as a side-product of a more fundamental deficit in perceiving artifacts. If Bayne were correct that the agnosic's deficit is primarily one of absent artifactual perception, then we might well expect this absence to have top-down effects on the holistic processing of presented artifacts, such that Bayne's explanation might predict that when presented with familiar objects, agnosics and control subjects will exhibit different patterns of integration of local spatial features. For instance, seeing some object as a pipe might have top-down influences on which parts of this object are attended, so if agnosics cannot see a pipe *as a pipe*, they might well exhibit atypical patterns of attention or of holistic processing when viewing pipes. But in the tasks described, the objects presented were novel, i.e., non-artifactual, and hence, the patient's worse performance on these tasks cannot be explained as the consequence of absent artifactual perception. Rather, the agnosic's performance on these tasks powerfully suggests a deficit in holistic processing that is independent of any purported deficit in artifactual perception.

Given that the described patient cannot simultaneously integrate an object's multiple spatial relations, we should expect him to exhibit difficulties in determining whether a visually-presented object is a stethoscope or a matchbook or a pipe, just as he struggles to distinguish between novel objects. That is to say, the deficit in holistic processing exhibited by this patient suffices to account for his categorization impairment. We need not further posit that he exhibits an abnormal absence of artifactual perception. On this view, what distinguishes the agnosic from this typical counterpart is *merely* a deficit in holistic processing: the agnosic has such a deficit, whereas his typical counterpart does not. Thus, we need not accept Bayne's view that what distinguishes the agnosic from his typical counterpart is that the agnosic cannot visually perceive objects as pipes

or as guitars, whereas his counterpart can visually perceive objects as pipes or as guitars. While it is true that the patient lacks the ability to perceive artifactual features, this does not distinguish him from his counterpart, who *also* does not perceive artifactual features.

The evidence for a holistic deficit in associative agnosia rests solely on Farah's (1990) review and on Delvenne's (2004) case study. Therefore, it may turn out that future cases of associative agnosia will be discovered which do not involve a deficit in holistic processing and which are therefore more plausibly explained in terms of a deficit in the perception of artifacts. However, given the current state of evidence, we have no reason for thinking such a form of agnosia exists and correlatively no reason for thinking that to explain agnosia, we must posit that typical perceivers visually perceive objects as artifacts.

6 Conclusion

In this essay, I have criticized three lines of argument in favor of high-level perception. These are: introspective arguments, Siegel's argument from felt unity, and Bayne's argument from agnosia. Against introspective methods, I have suggested that introspection on the target experience is neutral on the question of whether these experiences are wholly perceptual, cognitive, or some combination of these. Against Siegel's argument, I have suggested that the feeling of unity that obtains between low-level visual information and certain high-level representations might be elicited by certain visual-cognitive configurations and thus, that felt unity provides no reason to favor a purely visual explanation of high-level experiences over its cognitive counterpart. Against Bayne's argument from agnosia, I have drawn from empirical evidence to argue that associative agnosia is best explained by a deficit in holistic processing, without further positing that agnosics have an abnormal absence of artifactual perception.

The arguments for high-level perception I have considered do not exhaust strategies in the literature; there are many other interesting and promising approaches to high-level perception.

For instance, Ned Block has recently argued that adaptation effects might be used as part of an argument in favor of some kinds of high-level perception (Block, 2014). William Fish has suggested that a range of empirical results, including response times, might be used to argue for certain kinds of high-level perception (Fish, 2012)³⁷ These arguments merit consideration elsewhere. I will merely note that if these arguments are to succeed, they must suggest a reason to think that fast, automatic, non-consciously processed post-perceptual judgments cannot account for the relevant phenomena.

In the next and final essay in the dissertation, “Visually Perceiving the Intentions of Others,” I defend an alternate route to the view that we perceive high-level features. This strategy begins with the claim that certain experiences which represent high-level features are rationally unresponsive to certain kinds of evidence. Given the view of belief I defend in “The Revisability View of Belief,” these experiences cannot be judgments, even of a fast, automatic, non-consciously processed kind. On the next best explanation, these experiences are perceptual. I use this strategy to argue in particular that we can sometimes perceive the intentions of others, but the general strategy might turn out to be usable to show that we enjoy other sorts of high-level perception as well.

³⁷ Fish (2012), following Block (2010), also supports the use of adaptation effects to argue for high-level perception.

Visually Perceiving the Intentions of Others

3

In a ground-breaking 1944 study, the psychologists Fritz Heider and Marianne Simmel presented subjects with a short film depicting a small triangle, a large triangle, and a circle, moving about a stationary, partially-filled in square. When asked to describe the events in the film, nearly all subjects produced narratives which ascribed intentions and desires to the two-dimensional shapes. For instance, in one scene of the film, the large triangle moves to the right of the screen, a short distance behind the small triangle and circle, which also move to the right. Subjects described this scene as one in which the large triangle *chases* the other shapes and one in which the smaller shapes *flee from* the triangle. In another scene, the small triangle and circle remain motionless within the stationary square, while the large triangle moves back and forth immediately outside the perimeter of the square. Subjects described this scene as one in which the smaller shapes *hide from* the large triangle (Heider and Simmel, 1944). Though subjects might have described these scenes in purely geometric and physical terms, their descriptions consistently treat the two-dimensional figures as alive, and as acting on intentions.

It is difficult for subjects to interpret the movie in terms that do not involve actions and aims. Even when subjects are explicitly instructed to provide a narrative in terms of purely physical, geometric terms, they struggle to do so (Hashimoto, 1966). Viewing the Heider and Simmel movie for oneself makes it clear why this is so. The figures just *seem* to be fleeing, hiding, chasing, and even fighting; this reading is effortless and immediate (you can watch the video here: https://www.youtube.com/watch?v=wp8ebj_yRI4).

These results together suggest that the typical subject viewing the Heider and Simmel movie *experiences* the two-dimensional shapes as fleeing, hiding, chasing, and so on, in some

sense of ‘experiences.’ The question remains what kind of experience this is, that is, what kind of mental attitude or attitudes make up this experience. To keep things suitably concrete, I will focus in particular on the scene in which the smaller shapes ‘flee’ the large triangle and will henceforth refer to this experience as *the ‘fleeing’ experience*. On one view, the subject viewing this scene in the Heider and Simmel movie enjoys a *visual perception* as of the smaller shapes fleeing the triangle (Scholl & Gao, 2013; Scholl & Tremoulet, 2000). On another view, the subject merely perceives shape, color, and movement features, and the perception of these features quickly and automatically generates a post-perceptual *thought* that the smaller shapes are fleeing the large triangle (Butterfill, 2009).³⁸

The question of whether perception can represent intentional actions is merely one instance of a more general question about whether perception can represent *high-level features*. Recently proposed high-level percepts include: artifactual features (e.g., *being a guitar*), first-personal agency, dispositional features, absence, event-causal features, natural kind features (e.g., *being a lemur*), a range of social features (e.g., *being a face*, *being masculine*, *being an expression of happiness*), and moral features (Audi, 2013; Bayne, 2008; Bayne, 2011; Block, 2014, Fish, 2012; Chappell, 2008; Cullison, 2010; Farennikova, 2013, Nanay, 2011; Butterfill, 2009; Scholl & Tremoulet, 2000; Siegel, 2005; Siegel, 2009; Sorensen, 2012; cf. Chalmers, 2006; Pautz, 2009; Prinz, 2013; Rips, 2011)

I will argue that the typical subject viewing the Heider and Simmel movie *visually perceives* the smaller shapes as fleeing the triangle. Moreover, in visually perceiving the smaller shapes as fleeing the triangle, the subject thereby perceives the smaller shapes as acting on a certain *proximal* intention, the intention of evading the large triangle. Thus, just as we can visually

³⁸ Shea (2014), following Carey (2009), places animacy in a borderline case, of a kind that is neither determinately cognitive nor determinately perceptual.

perceive something as blue or as trapezoidal or as moving to the left, so too can we visually perceive the intentions of others.

The argument I offer in favor of the view that we visually perceive others' intentions is rooted in the claim that the 'fleeing' experience is in a certain sense unrevisable, viz., it lacks a nomic capacity to be rationally revised in response to any bit of relevant evidence that conflicts with it. I consider and criticize the following three cognitive explanations of the 'fleeing' experience:

- the 'fleeing' experience is a *judgment*
- the 'fleeing' experience is a *System 1 output*
- the 'fleeing' experience is a *delusion-like cognitive attitude*.

I argue that none of these explanations can account for the 'fleeing' experience's unrevisability. In contrast, if the 'fleeing' experience is a visual experience, then we can account for the experience's unrevisability by explaining it in terms of the mechanisms that underlie more familiar visual illusions, such as the Müller-Lyer Illusion and Titchener's Illusion.

The paper is structured as follows: In §1, I clarify the essay's thesis and suggest why it matters whether it is true. In §2, I argue that the visual explanation of the 'fleeing' experience is *prima facie* plausible and is resilient enough to survive certain initial worries. In §3, I argue that the doxastic explanation, suitably understood, is also *prima facie* plausible and resilient enough to survive certain initial worries. In §4, I argue that the fact that the 'fleeing' experience is *judgment-discordant*, or can occur at the same time as relevant conflicting judgments, provides no reason to think that the 'fleeing' experience is not itself a judgment. In §5, I argue that the 'fleeing' experience lacks a certain capacity for rational revision and is therefore not a judgment. In §6, I argue that the 'fleeing' experience is not a System 1 output. In §7, I argue that the 'fleeing' experience is also not a delusion-like cognitive attitude. In §8, I conclude in favor of the visual explanation of the 'fleeing' experience.

0 Introduction

First, a terminological point. I have already been using ‘experience’ in a somewhat non-standard way, as an attitude-neutral term for any bit of consciousness. On this way of talking, perceptual states count as experiences, as do judgments, feelings, and desires. Experiences can further comprise multiple, distinct attitude types. For instance *the experience of seeing an apple and desiring to eat it* is composed of two different attitudes, a perceptual state and a desire. The reason we need an attitude-neutral mechanism of talking about these clusters of mental states is that the major thesis of the paper is that intentions are perceived, and not merely represented by a combination of perception and cognition. Hence, we require a way of describing this bit of consciousness without presupposing which sort of attitude or attitudes make it up.

The locution ‘perception *as of* Φ ’ is generally non-committal as to whether the relevant perceptual state is accurate, and the locution ‘perception *of* Φ ’ is sometimes reserved for accurate perception. I generally adhere to this convention, but occasionally abandon it for the sake of naturalness. Throughout, all talk of perception should be understood to be talk of ‘as of’ perception.

Throughout, I also talk as though features are constituents of experience. I adopt this way of talking for the sake of naturalness, and I am neutral on the issue of whether perceptual content is exhausted by states of affairs, instead of by a Fregean sense, or an intension. I am further neutral on whether the contents of perception are structured propositionally, attributively, iconically, or otherwise.

Finally, I am making at the outset a substantive assumption about perceptual content, that it is *naïve* or non-phenomenal, in the sense that perceptual experiences at least sometimes ascribe mind-independent features to objects, such as *being spherical* or *being grooved*. The naïve approach to perceptual content contrasts with the phenomenal approach, on which perception only

ever ascribes mind-dependent features to objects, such as *looking spherical* or *feeling grooved* (Glüer, 2009).

1 The Thesis and Why It Matters

The thesis of this paper is that we sometimes visually perceive the intentions of others. This thesis requires clarification on a number of dimensions. First, there is an important question of what is meant by *visual perception*. In the relevant sense, visual perception is the kind of environmentally-originated experience whose paradigm instances include the visual perception as of something blue or the visual perception as of something spherical.

Visual perception in this sense does not extend to certain more extended senses of ‘seeing.’ For instance, there are more extended senses of ‘seeing’ on which someone who has excellent evidence about who will win an upcoming election can see who the winner will be, or on which someone viewing a room in disarray can see that it will take several hours to clean up. The claim that we visually perceive others’ intentions is meant to be true in the more restricted sense of seeing whose paradigm instances include the visual perception as of something blue. Should it turn out to be the case that the only sense in which we see the intentions of others is in the extended sense of ‘seeing,’ the present thesis would not turn out to be true.

The claim that we visually perceive the intentions of others can further be precisified as the claim that we sometimes enjoy visual experiences *as of* the intentions of others. These experiences might be inaccurate, as when the ascribed intention is not in fact possessed by the entity it is ascribed to. Just as one might experience a red object *as* blue or a two-dimensional object *as* three-dimensional, so too might one experience someone who lacks a certain intention *as* bearing that intention, even when that person lacks the relevant intention.

Finally, the thesis that we perceive the intentions of others is the claim that *typical* humans sometimes visually perceive the intentions of others.³⁹ That typical humans enjoy this capacity is a highly contingent fact about them. Such humans might have been otherwise, as demonstrated by the fact that many humans *are* otherwise: subjects who are autistic or who have lesions of the amygdala do not visually perceive the intentions of others (Congiu et al., 2010; Heberlein & Adolphs, 2004). Analogously, it's an interesting fact about typical humans that they visually perceive movement, even though this is a contingent capacity, as demonstrated by the fact that some humans, viz. those with *akinetopsia*, do not see movement.

Finally, the thesis that we visually perceive the intentions of others does not entail the stronger claim that the perception of others' intentions is unique to the visual modality. There is a case to be made that the intentions of others can be represented by other modalities, such as the auditory modality. Consider that one difference between autistic and non-autistic subjects is that non-autistic subjects are better at detecting sarcasm than their autistic counterparts (Rutherford, Baron-Cohen, & Wainwright, 2002). Arguably, this difference is explained by a difference in the auditory perception of a communicative intention. If this is so, then others' intentions might be something that some humans can hear, as well as see. Here too the perception of others' intentions is analogous to the perception of movement, which can occur in a variety of modalities.

One reason the particular issue of whether we can perceive intentions matters is that it may help illuminate certain issues in psychopathology. For instance, we know that certain *disorders of theory of mind*—including autism spectrum disorders—involve deficits in attributing intentions (Congiu et al., 2010; Heberlein and Adolphs, 2004). Whether these deficits are at root

³⁹ This claim is emphatically not the value-laden claim that *normal* humans sometimes see the intentions of others. It is merely the descriptive claim that a statistical majority of humans visually perceive the intentions of others.

perceptual or whether they are merely cognitive remains a major lacuna in our understanding of these disorders.

A second reason it matters whether we can visually perceive others' intentions is that if we can, this would help clear ground for a *direct perceptual view* of knowledge of other minds, on which we come to know about others' mental states through direct perception, instead of by applying a theory or by simulating their experiences (Gallagher, 2008). Conversely, if do not visually perceive others' intentions, the direct perceptual view of knowledge of other minds cannot get off the ground.

In the next two sections, I argue that both the visual and doxastic explanations of the 'fleeing' experience are resilient enough to survive certain initial objections. I post-pone until sections 6 and 7 a discussion of other cognitive explanations of the 'fleeing' experience.

2 The Initial Plausibility of the Visual Explanation

To appreciate the difference between the visual and doxastic explanations of the 'fleeing' experience, notice that the claim that we see the intentions of others is comprised of a *content* claim and an *attitude* claim. The visual and doxastic explanations are in broad agreement about the content claim: they both posit that the smaller shapes are represented as fleeing the large triangle. The visual and doxastic explanations differ only on the matter of attitude: the visual explanation posits that the 'fleeing' experience occurs at the level of visual experience, whereas the doxastic explanation posits that this experience is merely a post-perceptual judgment.

Before turning to a further comparison of the visual and doxastic explanations, it is important to appreciate why we ought to accept the content claim. Why think that intentions are represented by the 'fleeing' experience? In the paper's introduction, I have already briefly sug-

gested some reasons to think this experience represents fleeing: subjects immediately and irresistibly describe the movie in terms of a fleeing narrative, and viewing the Heider and Simmel movie for oneself elicits a strong impression of fleeing.

I will suppose going forward that the typical subject viewing the Heider and Simmel move represents, *somehow or other*, the smaller shapes as fleeing the large shapes. The question remains why we should suppose this to necessarily involve attributing to the smaller shapes any particular intention. As it turns out, fleeing constitutively involves a proximal intention, the proximal intention of evasion. For this reason, to experience someone as fleeing is thereby to experience them as acting on the intention of evasion.

In general, some entity A flees some entity B only if A self-propels with the proximal aim of evading B. *Self-propulsion* is any movement of the self which is facilitated by or guided by the self. Self-propulsion can occur via biological movements, such as running, crawling, swimming, or slithering. But it can also occur via assisted movements. I can self-propel by riding a motorcycle, by paddling a kayak, or by asking someone to push me while I sit in a wheelbarrow. The creature who cannot self-propel cannot flee, even if that creature has the aim of evading something or someone else. Such an entity cannot act on this aim, which is what fleeing requires.

Fleeing necessarily entails a *proximal intention* of evasion, though not necessarily a *distal intention* of evasion (Mele & Moser, 1994). The distinction between proximal and distal intentions can be illustrated by an example. My six-year-old nephew sometimes catches my eye, smiles, and then runs away from me. He knows that when he does this, I'm likely to chase him, catch him, and tickle him. In running away from me, my nephew acts on the distal intention of getting caught, since his ultimate intention is to get me to catch him and tickle him. But in fleeing, he acts on the proximal intention of evading me and my tickling hands. My nephew's distal and proximal intentions in fleeing thus come apart.

That a proximal intention of evasion is necessary for some act of self-propulsion to count as fleeing can be illustrated by an example: if you go on a jog in a wooded area and, unbeknownst to you, a curious squirrel scampers after you, you do not flee the squirrel, even though you do self-propel away from the squirrel. Plausibly, the reason your self-propulsion away from the squirrel does not count as an act of fleeing the squirrel is that you lack a proximal intention of evading the squirrel. This suggests a reason to think that this sort of proximal intention is a necessary condition on fleeing.

The act of fleeing is often grounded in a further, more distal aim, such as avoiding entrapment, winning a game, distracting the fleeing party, helping someone exercise, or even—as in the example of my nephew—getting caught. While it is at least in principle possible that we might sometimes perceive others' distal intentions, the present thesis is restricted to the more modest claim that we sometimes visually perceive proximal intentions, and in particular those proximal intentions that one evinces by certain actions.

Going forward, then, I will assume that the experience as of the smaller shapes fleeing the triangle is partly an experience as of the smaller shapes proximally intending to evade the triangle. So, if we see the smaller shapes as fleeing the triangle, we also see them as proximally intending to evade the triangle.

An initial worry about the visual explanation stems from the fact that the nature of intentions is a fraught theoretical issue. Several important issues divide theorists: whether intentions are necessarily represented, whether unforeseen consequences can count as intentional, whether intentions are causally efficacious in action, and whether intentions bear constitutive connections to judgment (Setiya, 2014).

Given the uncertainty about what intentions are, it may seem unlikely that we might adduce much evidence in favor of the view that we see intentions. There are two worries in this

vicinity. The first is the methodological concern that we will not be able to settle whether we see intentions before we settle on the best theory of intentions. The second is the concern that if intentions turn out to be of a sophisticated kind, it is implausible that they might be perceived, since it is implausible that we might perceive something of such an arcane nature.

Both of these concerns can be addressed simultaneously: if we perceive intentions at all, we perceive them without also perceiving their deep nature. The perception of intentions is simply silent about the deep nature of intentions. And this is unsurprising: in general, perception as of Φ does not typically (if ever) involve the perception as of the deep nature of Φ . This point is cogently argued by Susanna Siegel, who makes it by reflecting on the perception of holes:

Suppose you see some cheese with a hole in it . . . According to one theory of the metaphysics of holes, they are immaterial particulars. According to an opposing theory, they are material but negative parts of material particulars. When the cheese appears to have a hole in it, does it appear to host an immaterial particular, or does it appear to have a material part – or neither, or both? . . . The best answer seems to be ‘neither’: visual experience just seems neutral on whether the hole in the cheese is a material but negative part or an immaterial part hosted by a material particular (Siegel, 2009, p. 539).

Just as we can perceive the hole in a slice of Swiss cheese without perceiving it as (say) a negative part of a material particular, so too can we perceive someone as proximally intending to evade someone or something else without thereby perceiving that agent as (say) representing such an intention. Our experience of others’ intentions is simply neutral on whether those intentions are necessarily represented, causally efficacious, and so on.⁴⁰

⁴⁰ The claim that we can see some object without thereby seeing its deeper nature further underscores that the relevant sense of seeing is *seeing as* and not, for instance, the ‘simple seeing’ that Dretske advocates (Dretske 1979).

3 The Initial Plausibility of the Doxastic Explanation

On the doxastic explanation of the ‘fleeing’ experience, this experience involves visual perception as of shape, color, and movement features, and this perceptual experience triggers a post-perceptual judgment ‘the smaller shapes are fleeing the triangle.’ Perception itself does not represent fleeing.

At first glance, the doxastic explanation might appear to be implausible. After all, the ‘fleeing’ experience is stimulus-driven, automatic, and irresistible. These features are more typically associated with perception than judgment. I want to suggest that there is a particular version of the doxastic explanation that can account for all of these aspects of the ‘fleeing’ experience. On this view, the post-perceptual judgment ‘the smaller shapes are fleeing the triangle’ is formed automatically, quickly, and—though itself conscious—is formed on the basis of processes which are non-conscious. There is broad empirical evidence for judgments of this sort. For instance, consider the following math problem; answer it as quickly as you can:

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? Answer: ____ minutes

If you are like most people, the answer *100* probably popped into your head (Thompson, 2009). However, the correct answer to the question is *5*. That most participants give an incorrect answer is somewhat surprising, given that the correct answer can be arrived at using a bit of simple algebra.⁴¹

⁴¹ The trick is to think of the rate at which the machines generate widgets. 5 machines work at a rate of 1 widget per minute. Therefore, 100 machines work at a rate of 20 widgets per minute, which means it will take 100 machines *5 minutes* to complete 100 widgets.

The standard explanation for the fact that the incorrect answer immediately comes to mind is that this answer is arrived at by something like a *matching heuristic*: the system notices the pattern 5:5:5 and the partially completed pattern 100:100: ____ and generates ‘100’ as the correct answer (Thompson, 2009).

I say that ‘the system’ notices the relevant pattern and not ‘the subject’ because it’s highly unlikely that the typical subject applies the matching heuristic by a process of explicit inference, or by conscious construction of an argument. This is for two reasons. First, inference is time-consuming, and the incorrect answer is produced very quickly. Second, had the subject engaged in inference, it’s unlikely she would have made the mistake of applying the matching heuristic to the ‘widgets’ problem, since a moment’s consideration shows the application of this heuristic to this problem to be inappropriate. The ‘widgets’ problem is clearly a rate task, and there is nothing in the semantics of the problem that might reasonably suggest it to be a mere matching task. Thus, on the best explanation, the attractive, incorrect judgment is the conscious output of a process which is itself non-conscious. Relatedly, the incorrect answer to the problem is not associated with effortful deliberation or a feeling of initiating some process. One simply ‘sees’ the answer.

The ‘widgets’ problem is just one of many problems which tend to elicit automatic, fast, and non-consciously processed judgments. In the literature on cognition, such outputs are commonly referred to as *System 1 outputs*, whereas slow, analytically and effortfully produced outputs are commonly referred to as *System 2 outputs* (Evans, 2008).⁴² In appealing to such judgments,

⁴² This terminology reflects the common presumption that System 1 outputs and System 2 outputs derive from functionally distinctive systems. Here, we can be neutral on whether dual systems underlie these outputs. All this is required for present purposes is that some outputs are quickly, heuristically, and automatically produced, whereas others are slowly, automatically, and effortfully produced.

the advocate of the doxastic explanation of the ‘fleeing’ experience enjoys broad independent support.

A variant of the doxastic explanation which appeals to a fast, automatic, non-consciously processed judgments to account for the ‘fleeing’ experience can accommodate the subjective quality of the ‘fleeing’ experience. According to such a view, it is because the relevant judgment is generated very quickly after the onset of the visual perception of shape and color that there is not any discernible temporal gap between the perception of shape and color and the experience as of the smaller shapes fleeing the triangle. Because the judgment is formed without initiation or effort, there is no feeling of effort, or awareness of initiating an inferential process. Finally, because the relevant judgment is formed on the basis of a non-conscious process, subjects enjoying the ‘fleeing’ experience do not experience themselves as inferring that the smaller shapes are fleeing the triangle. In sum, the cognitive explanation is well-positioned to account for the fact that it just *seems* as though the smaller shapes are fleeing the triangle.

In this section, I have argued that both the perceptual and doxastic explanations of the ‘fleeing’ experience are resilient enough to handle certain initial objections. In the next section, I consider an argument which some have offered against the doxastic explanation. This is the *argument from judgment-discordance*. I will suggest that this argument ultimately provides no reason to endorse the perceptual view over its doxastic competitor.

4 Against the Argument from Judgment-discordance

Given that both the perceptual and doxastic explanations of the ‘fleeing’ experience are prima facie plausible, how do we adjudicate between them? One argument attempts to exploit the fact that the experience of ‘fleeing’ is *judgment-discordant*, or can persist at the very same time as a judgment which conflicts with it. The fact that a mental state exhibits judgment-discordance has sometimes been used to argue that that mental state is not a judgment. For instance, to argue

against a doxastic explanation of ‘pine tree phenomenology,’ Susanna Siegel argues that this ‘pine tree phenomenology’ might persist even if one were to ‘cease to dwell on the belief that you’re looking at a familiar tree:’

Suppose that you’re an expert pine-spotter looking at some pine trees in the forest. Then someone tells you that the forest has been replaced by an elaborate hologram, causing you to cease to dwell on the belief that you’re looking at a familiar tree. If an event such as [dwelling on the belief *that* is a familiar tree] were what contributed to the phenomenological change before and after your acquiring the disposition to recognize pine trees, then we would expect your acceptance of the hologram story to make the hologram look as the forest looked to you before you knew how to recognize pine trees. But, intuitively, the hologram could look exactly the same as the forest looked to you after you became an expert. So the familiarity with pine trees does not seem to have its phenomenological effects at the level of belief (Siegel, 2010, p. 104-5).

This argument contains an implicit assumption: that if you ‘cease to dwell on the belief that you’re looking at a familiar tree,’ then you don’t simultaneously dwell on the belief or judgment that you’re looking at a familiar tree. On this assumption, judgments never or rarely simultaneously conflict. This style of argument might be generalized as follows: if some experience as of p can exist at the same time as some explicit judgment not- p , then the experience as of p is not itself a judgment, since explicit judgments cannot—or usually do not—conflict. Call this general argument the argument from judgment-discordance.⁴³

⁴³ Another recent example of arguing in favor of a certain kind of high-level perception by appealing to judgment-discordance comes from Bayne. To argue in favor of the perception of first-personal agency, Bayne considers the subject with utilization syndrome and argues:

one might convince the patient with utilization behavior that she is not acting despite the fact that she experiences herself as acting ... A patient simultaneously manifesting an anarchic hand and utilization behaviour might believe that both (or neither) of the movements in question constitute action of theirs, but none of this will change his or her agentive experience (Bayne, 2006, p. 186).

One might use the argument from judgment-discordance to attempt to show that the ‘fleeing’ experience is not a judgment, as follows: the typical subject viewing the Heider and Simmel movie fully appreciates that the shapes she is viewing are inanimate figures, incapable of acting or intending. She might even explicitly judge ‘the shapes are not *really* fleeing the triangle.’ Nevertheless, her experience as of the shapes fleeing the triangle remains. Since explicit judgments never—or only rarely—conflict, it is impossible or unlikely that her experience as of the shapes fleeing the triangle is itself a judgment. The doxastic explanation of the ‘fleeing’ experience is thus a highly implausible one.

Before discussing the argument from judgment-discordance further, it may be helpful to rule out a certain preliminary objection to the style of argument, one which arises from the fact that the world of the inanimate figures is arguably a *fictional* world. Evaluated in the context of the relevant fictional world, it is arguably true that the smaller shapes flee the triangle, just as, evaluated relative to the fictional world of *Pride and Prejudice*, it is arguably true that Ms. Bennett wishes to marry Mr. Darcy. If the experience as of the smaller shapes fleeing the triangle is evaluated relative to the fictional world which these shapes inhabit, and if the judgment ‘it’s not the case that the smaller shapes are fleeing the triangle’ is evaluated relative to the actual world, then perhaps these attitudes are not in conflict, since they are evaluated relative to different contexts.

It’s quite plausible that the shapes of the Heider and Simmel movie inhabit a fictional world. However, it’s important to see that this aspect of the movie—that it depicts a fictional

Bayne implicitly assumes that if the atypical agentive experience could persist at the same time as a relevant conflicting judgment, then the experience cannot itself be a judgment.

world—is not necessary for illusions of intentionality. Rather, illusions of animacy and of intentions can arise even in contexts that do not involve fiction or fantasy. For instance, consider a possible paradigm involving three inanimate, plastic objects—a large pyramid, a small pyramid, and a sphere. Suppose further that these objects are remotely controlled by a researcher to create a narrative that is roughly isomorphic to the narrative depicted by the Heider and Simmel movie. It seems extremely likely that the three-dimensional version of the Heider and Simmel movie would trigger a ‘fleeing’ experience of the kind triggered by the two-dimensional Heider and Simmel movie. But it’s less obvious that the three-dimensional analog of the Heider and Simmel movie describes a fictional world. Since we can causally interact with the remote-controlled plastic shapes, it would seem that they occupy our world, and not a fictional one.

In the three-dimensional version of the Heider and Simmel paradigm, it would not be possible to appeal to different contexts of evaluation to account for the fact that the subject judges ‘it’s not the case that the smaller objects are fleeing the pyramid’ and simultaneously experiences the small objects as fleeing the pyramid. In this version of the case, there is no fictional, further context of evaluation that might explain away any apparent conflict between the subject’s judgment and her experience.

Going forward, I will continue to talk in terms of the two-dimensional version of the Heider and Simmel paradigm, since this is the version of the paradigm that has been most empirically investigated. But I will suppose that the fact that the Heider and Simmel movie describes a fictional world does not explain why it gives rise to an illusion of intentions. For this reason I will also suppose that the conflict between the subject’s ‘fleeing’ experience and the subject’s judgment

that the small shapes are not fleeing the triangle cannot be explained away by appealing to different contexts of evaluation.⁴⁴

Having briefly explained why I am not moved by the ‘fictional worlds’ objection to the argument from judgment-discordance, I want to suggest what does motivate my rejection of the argument. I reject the premise of the argument that explicit judgments never—or perhaps only rarely—conflict. As several theorists have by now pointed out, judgments can—and frequently do—conflict, even when both judgments are each held to a high degree, attended, overtly contradictory, and explicitly held (Byrne, 2009, pp. 450-1; Gertler, 2011; Huddleston, 2012; Mylopoulos, 2015, p. 768). Here I focus on two such cases: cognitive illusions and superstitions.

Suppose you are a contestant on a game show in which you can choose from three closed doors for a chance of a prize. One door conceals the prize, while the other two conceal worthless

⁴⁴ One might object that even the three-dimensional version of the animacy paradigm might depict a fictional world, just as (real, three-dimensional) sock puppets might be used to depict a fictional narrative. A different case suggests that illusions of others’ intentions can arise even in the absence of fictional contexts: suppose you are watching two children, one on the outside of a glass sliding door and one on the inside of glass sliding door facing each other, apparently engaged in a game of mimicry. The child on the inside raises her right hand, and the child on the outside raises her left hand. The child on the inside raises her left foot, and the child on the right raises her right foot, and so on. Suppose that you come to learn that the glass door separating the children is entirely opaque and sound-proof (you are entirely sure on this point). What appears to be a game of miming is merely an incredible coincidence, as the children are not even aware of each other’s presence.

In such a case, you might nevertheless continue to experience the children as intending to imitate each other, even though you know there is no such intention. I take this case—though fanciful—to suggest that the presence of a fictional context is not necessary for the experience of intentions. The children’s actions are not part of a fictional narrative, but their behavior nevertheless elicits an intentions attribution.

objects. You choose one door but do not open it. The host of the show then selects one of the other two doors and opens it, revealing one of the worthless objects. From watching the show, you anticipated this move. At this point in the game, the host always opens a different door, and this door always reveals one of worthless objects. You now have a choice: you can choose to open the door you originally selected, or you can switch to the remaining closed door. What should you do?

It turns out that you should switch to the other closed door (Rosenhouse, 2009). This solution to the problem is extremely counter-intuitive. When the writer Marilyn vos Savant presented the Monty Hall problem along with its correct solution in a popular column, she received thousands of letters insisting that her solution was incorrect, many of them from professional mathematicians (Rosenhouse, 2009). Apparently, many subjects who entertain the Monty Hall problem initially incorrectly conclude that it does not matter whether you switch doors.

Consider one such subject, who is convinced by the incorrect answer to the problem. Suppose that, after thinking through the relevant conditional probabilities, this subject comes to judge that in fact, you should switch doors. Given how powerful her initial judgment is, she might not be able to shake it, at least not right away, even after coming to embrace the correct solution. If this is at all plausible, then there is some period, however brief, during which this subject enjoys two transparently contradictory beliefs: ‘you should switch to the other door’ and ‘it doesn’t matter whether you switch to the other door.’⁴⁵

The second kind of case that plausibly involves contradictory judgments is that of superstitious judgments held by subjects who know them to be false. Even subjects who are otherwise quite rational can harbor minor superstitions. Plausibly, some of these subjects hold these superstitions at the very same time they know them to be false. For instance, suppose you are scheduled

⁴⁵ I am especially indebted to David Chalmers for encouraging me to take seriously the possibility that cognitive illusions can engender contradictory judgments.

to give a job talk for your dream job in philosophy. In deciding what to wear for the talk, you choose your ‘lucky’ pair of socks. Of course, you know they’re not *really* lucky. While putting them on, you might even think to yourself how silly it is that you’ve chosen this pair. But you put them on anyway.

To take another case, suppose your loved one is awaiting the result of a biopsy test that will determine whether he has cancer. During this stressful period, you develop a superstition that if you don’t say aloud the word ‘cancer,’ the test result will be negative. You know full well that this is an irrational view, one you created to give yourself a sense of control during a difficult situation, but you nevertheless carefully abstain from saying the terrible word. On several occasions, you explicitly think to yourself how silly this superstition is and try to talk yourself out of it. But the superstition persists (Huddleston, 2012).

If cognitive illusions or rejected superstitions involve overtly contradictory judgments, it is not ad hoc to posit such judgments to explain the ‘fleeing’ experience. Therefore, the fact that the ‘fleeing’ experience can occur at the same time as a strongly held judgment that it is false does not tell against the doxastic explanation of the ‘fleeing’ experience.

Though the argument from judgment-discordance does not rule out the cognitive explanation of the ‘fleeing’ experience, the argument is nevertheless on the right track. In the next section, I sketch a different argument against the doxastic explanation of the ‘fleeing’ experience, one which is spiritually akin to the argument from judgment-discordance but ultimately stronger, in that it is usable in a narrower range of cases than the argument from judgment-discordance. The starting point of this argument is the claim that the ‘fleeing’ experience is not merely such that it can exist at the same time as a relevant conflicting judgment. The ‘fleeing’ experience is further *never* rationally revised in response to conflicting evidence, as evidenced by the fact that the ex-

perience is robust across subjects and across time. Thus, the ‘fleeing’ experience is *rationally immune* to at least some kinds of conflicting evidence, and this kind of evidence-immunity suggests a strong reason to think that ‘fleeing’ experience is not a judgment.

5 Against the Doxastic Explanation of the ‘Fleeing’ Experience

The argument against the doxastic explanation of the ‘fleeing’ experience is rooted in a view of belief (and judgment) I defend elsewhere. Before presenting the argument, it is necessary to sketch that view.

5.1 *The Revisability View of Belief*

On the *revisability view of belief*—which extends to doxastic states more generally, including judgments—all judgments, insofar as they are judgments, have a certain nomic capacity to be rationally revised in response to any piece of sufficiently strong, available conflicting evidence. This view requires clarification on a number of dimensions.

First, some mental state tokened in some creature is *nominally capable* of being rationally revised just in case, in at least some worlds in which the relevant creature’s psychological mechanisms are held fixed, that mental state is revised in response to evidence that conflicts with it. Second, some mental state is rationally revised in response to evidence that conflicts with it only if that mental state is revised: (i) in the right direction, (ii) in response to evidence, (iii) *and* via a non-deviant route. I will take these three conditions on rational revision in turn.

First, for some mental state to be rationally revised, it must be revised *in the right direction*. Which direction is the right direction is dictated by the evidence. For instance, if the best evidence suggests that not- p , then for the mental state as of p to be rationally revised, it must decrease in strength, disappear altogether, or be suspended. Should this mental state increase in strength, it is revised in the wrong direction and hence, the revision is not rational.

Second, for some mental state to be rationally revised, it must be revised in response to *evidence*. If you judge that today is Wednesday and then, as the result of an unfortunate encounter with lightning, lose this judgment, your judgment has been revised, but not on account of evidence. This route to judgment revision is non-evidential and hence, *arational*.

Finally, for some mental state to be rationally revised, it must be revised *via a non-deviant route*. The task of distinguishing deviant from non-deviant routes of rational revision is a very large one; indeed, it is a primary project of contemporary epistemology. For our purposes, it is sufficient to distinguish deviant from non-deviant routes ostensively, by a pair of contrastive cases.

Consider a case in which you believe that extra-terrestrials do not exist. You then read in a reliable newspaper that the government has captured one. On the basis of the report, you give up your belief that extra-terrestrials do not exist. This is a non-deviant route to revision. Compare this case to one in which you read in the newspaper that the government has captured an extra-terrestrial, and the shock causes you to fall out of your chair and bang your head. By sheer coincidence, the blow obliterates your belief that extra-terrestrials do not exist. In this case, you revise your belief in response to evidence but via a deviant route.

This concludes my explanation of what it is for some state to be rationally revised in response to evidence. There is another aspect of the revisability view of belief that is worth emphasizing: on this view, any mental state that is a judgment must be capable of being revised in response to *any* bit of available evidence that conflicts with it. Put slightly more formally, the relevant view says: for all x such that x is a judgment, and for all y such that y is some piece of sufficiently strong, available evidence that conflicts with x : x is nomically capable of being revised in response to y .

Importantly, the capacity to be revised in response to any piece of relevant evidence does

not entail a capacity to be revised simultaneously in response to all pieces of evidence. It may be that some judgment can be revised in response to evidence that p and can also be revised in response to some other piece of evidence that $not-p$ even though that belief cannot be revised simultaneously in response to both pieces of evidence. This restriction is consonant with how we think of other capacities: that Janelle can swim a mile and can play the *Rhapsody in Blue* clarinet solo does not suggest that Janelle can swim a mile *while* playing the *Rhapsody* solo.

The motivation for thinking that judgments, insofar as they *are* judgments, are nomically capable of being revised in response to evidence which conflicts with them is that two independently plausible principles together entail this result. Put roughly, these are: first, that all judgments, insofar as they are judgments, are rationally required to be revised in response to any bit of sufficiently strong, available evidence that conflicts with them; and second, that if some mental state is rationally required to be revised revise in response to some bit of evidence, then that state is nomically capable of being so revised.

A thorough defense of both of these claims is beyond the scope of the present project, but what is important for present purposes is that the motivation for the revisability view of judgment is independent of the question of how to distinguish perception from judgment. Going forward, I will presume that the foregoing view of judgment is correct. Readers who do not wish to follow me in this assumption are free to recast the argument as an argument for the conditional claim: *if* the relevant view of judgment is true, *then* we have good reason to think the ‘fleeing’ experience is not a post-perceptual judgment.

The relevant view of judgment imposes the following requirement on a doxastic explanation of the ‘fleeing’ experience: if the ‘fleeing’ experience is itself a judgment, then it must be revisable in response to any bit of evidence that conflicts with it. This requirement is consistent with the experience’s not *in fact* being revised, so long as there are circumstances in which it would be

revised. The remainder of the argument establishes that the ‘fleeing’ experience cannot meet this requirement and hence, is not a judgment.

5.2 The Argument from Non-revisability

Having briefly sketched the revisability view of judgment, I will now present the argument that the ‘fleeing’ experience is not a judgment. Let *S* be some typical subject *S* who views the Heider and Simmel movie under normal viewing conditions. Stated in terms of *S*, the argument is as follows:

- (1) *S* experiences the smaller shapes as fleeing the triangle.
- (2) If *S*’s experience as of the smaller shapes fleeing the triangle is a judgment, then that experience must be nomically capable of being rationally revised in response to any bit of sufficiently strong, available evidence that conflicts with it.
- (3) On the best evidence, *S*’s experience as of the smaller shapes fleeing the triangle is not nomically capable of being rationally revised in response to any bit of sufficiently strong, available evidence that conflicts with it.
- (4) On the best explanation, *S*’s experience as of the smaller shapes fleeing the triangle is not a post-perceptual judgment.

In broad strokes, the argument says that the ‘fleeing’ experience is unresponsive to conflicting evidence and that this fact rules out that it might be a judgment.

The first premise states *S* experiences the smaller shapes as fleeing the triangle. As throughout, ‘experience’ is neutral as to attitude, so this premise is a claim about *content*; it does not beg the question about whether perception is the vehicle of that content. The reasons for thinking that the typical subject viewing the Heider and Simmel movie represents the smaller shapes as fleeing the triangle have already been enumerated in the introduction to the paper: most subjects narrate the film in terms of fleeing; this interpretation of the film is difficult to resist; and

introspection of one's own experience reveals that it 'just seems' that the smaller shapes are fleeing the triangle.

The second premise states: *If S's experience as of the smaller shapes fleeing the triangle is a judgment, then that experience must be nomically capable of being rationally revised in response to any bit of sufficiently strong, available evidence that conflicts with it.* This premise just is an instance of the revisability view of belief. The revisability view says that all judgments (and beliefs) insofar as they are judgments, must be nomically capable of being revised in response to relevant conflicting evidence. Applying this view to the 'fleeing' experience, if this experience is a post-perceptual judgment, then this experience must be nomically capable of being revised in response to available, sufficiently strong evidence that conflicts with it.

The third premise states: *On the best evidence, S's experience as of the smaller shapes fleeing the triangle is not nomically capable of being rationally revised in response to any bit of sufficiently strong, available evidence that conflicts with it.* There are a couple reasons to think that subjects enjoying the 'fleeing' experience sometimes simultaneously token the judgment 'the smaller shapes are not fleeing the triangle.' The first is that subjects viewing the movie for the first time frequently experience *surprise* when viewing it. I suspect that this feeling arises because such subjects are aware that what they are experiencing is illusory. It's *because* subjects judge 'the smaller shapes aren't (really) fleeing the triangle' at the very same time that they experience the smaller shapes as fleeing the triangle that the experience surprises them. If subjects did not currently judge that the smaller shapes are not fleeing the triangle, it would be difficult to explain why subjects find the movie to be expectation-violating. They should find it no more surprising than seeing a blade of grass as green or than feeling an unsanded bit of wood as rough.

Relatedly, many subjects find the movie humorous. I once was present when the movie was shown to a classroom full of undergraduates, and the students laughed at several points in

the movie (I'm sure I did too, even though I'd seen it before). This is a common response. I suspect that part of the reason the movie is amusing is *that* it is incongruent with better judgment. One knows the relevant shapes are merely inanimate figures, but the drama comes to life anyway. Going forward, I will suppose that it is not only possible—but also quite common—for subjects enjoying the 'fleeing' experience to simultaneously judge that the smaller shapes are not fleeing the triangle.

Given that most or at least many subjects viewing the Heider and Simmel movie simultaneously judge their experience to be illusory, and given that this judgment is supported by good evidence, these subjects enjoy very good evidence that their experience is inaccurate. Nevertheless, their experience is not revised in response to this conflicting evidence. If it did, subjects' 'fleeing' experience would weaken in strength or disappear altogether. The figures would then appear to be mere shapes moving in geometrically describable ways, and not agents performing actions in a fluidly interpersonal fashion.

Not only does the 'fleeing' experience sometimes fail to be revised in response to the evidence that the figures it describes are not really acting on intentions, it appears to altogether lack the nomic capacity to be so revised in response to this evidence. This claim is much stronger than the previously discussed thesis that the 'fleeing' experience merely can persist at the same time as a conflicting judgment. On this claim, in all worlds in which the subject's psychological mechanisms are held fixed, the 'fleeing' experience is not rationally revised.

There are two main sources of evidence that the 'fleeing' experience is rationally immune to the judgment that the figures are not really acting on intentions. First, the 'fleeing' experience reliably occurs for most subjects. Indeed, the only subjects who do not experience it are those who have an autism spectrum disorder or lesions of the amygdala (Scholl & Gao, 2013). If this experience could be revised—for instance, if increased attention or deeper consideration would bring

about its revision—we would expect at least some typical subjects to lack or lose the ‘fleeing’ experience. For instance, we would expect that subjects with greater working memory or a greater tendency to engage analytic thinking would lack altogether or come to lose the ‘fleeing’ experience, but we don’t find any such results.

Second, it might be suggested that the reason subjects’ ‘fleeing’ experience persists despite the evidence against it is that subjects lack the requisite motivation to revise their experience. On one way of spelling out this worry, subjects can revise their ‘fleeing’ experience in response to the evidence, but they aren’t so inclined because, having been raised on cartoons, movies, and other fictional two-dimensional displays, the typical subject tested on this paradigm has acquired a delight for animate fictions and is not disposed to consider the rational import of these fantasies. On this hypothesis, the persistence of the ‘fleeing’ experience is a culturally-specific result seen only in those subjects who have been trained to indulge the pleasures of fantasy over reality.

The view that the ‘fleeing’ experience is nomically revisable, but that subjects’ culturally-entrained delight in fiction obliterates the motivation necessary for revision, is refuted by evidence that illusions of others’ intentions can be reliably elicited in subjects whose cultures lack cartoons and other animated fictions. In particular, in one study, animated figures elicited intentions attributions equally in Germans and in hunter-horticulturist Shuar subjects (Barrett, Todd, Miller, & Blythe, 2005).

A further reply to the claim that the ‘fleeing’ experience is unrevisable might be that subjects viewing the Heider and Simmel movie fail to internalize the evidence that the figures are not really acting on intentions. Internalizing new evidence can take time, so it might be thought that the ‘fleeing’ experience would be rationally revised, if sufficient time were permitted. The problem with this thesis is that there is evidence the ‘fleeing’ experience remains robust for subjects even after typical viewings. For instance, researchers working on the perception of animacy, who have

seen this stimulus many times, report that it continues to elicit the characteristic animacy interpretation even after multiple viewings (Scholl & Gao, 2013).

That the ‘fleeing’ experience is stable across subjects, across cultures, and across time does not entail that there are no circumstances in which the ‘fleeing’ experience might be destroyed or attenuated. There are probably many such circumstances, but these circumstances don’t obviously involve *rational* revision of the ‘fleeing’ experience. For instance, suppose it should turn out to be possible to destroy the ‘fleeing’ experience by means of broadly Pavlovian, associative process, for instance by administering an electrical shock every time a subject enjoys the ‘fleeing’ interpretation of the figures as opposed to a merely ‘geometric’ interpretation of them. This kind of brainwashing is not evidence-based, so it cannot count as a rational revision of the kind judgments must be capable of undergoing. Likewise, if one should desire to have experiences that are accurate and, moved by this desire, train oneself by a similarly associative process to ‘unsee’ the ‘fleeing’ experience, this would also not be a rational revision. Such a revision is made because of evidence, but occurs via a non-deviant route.

That the ‘fleeing’ experience is not rationally revised in cases for which we have evidence does not rule out a further possibility, on which the ‘fleeing’ experience is nomically revisable, but this capacity is for some reason *masked* in cases for which we have evidence. Just as a glass’ capacity to break might be masked in worlds in which it is wrapped in soft, thick padding, so too might the ‘fleeing’ experience’s capacity to be revised be masked by contingent features. These masking features do not obliterate the ‘fleeing’ experience’s capacity for revision; they merely obscure it (Bird, 1998; Johnston, 1992).⁴⁶

⁴⁶ This characterization of masks borrows from the dispositions literature, which is the context in which Bird and Johnston discuss masks.

What are we to make of the suggestion that the ‘fleeing’ experience is nomically capable of being revised in response to conflicting judgments, but that it simply isn’t revised in those circumstances for which we have evidence? As a theoretical possibility, this claim cannot be altogether ruled out. After all, for *all we know*, in some of the circumstances for which we don’t have evidence, the ‘fleeing’ experience is revised in response to the judgment that conflicts with it. Though the possibility that the ‘fleeing’ experience is revised in certain yet to be observed circumstances cannot be ruled out, we should like some reason to think the ‘fleeing’ experience will be revised in other circumstances before we take this possibility seriously. *What* would have to change in order for the ‘fleeing’ experience to become evidentially responsive? Do subjects need more time to reflect on the fact that the experience is illusory? Would better attention or increased effort help? We presently have no reason to think any of these circumstances would result in the ‘fleeing’ experience’s being rationally revised.

If all judgments are such that, insofar as they are judgments, they must be capable of being revised in response to conflicting evidence, and if the best evidence suggests that the ‘fleeing’ experience lacks this capacity, we have strong grounds for concluding that the ‘fleeing’ experience is not a judgment. This conclusion does not rule out that the ‘fleeing’ experience is some other sort of cognitive attitude. I devote the next two sections to excluding other cognitive explanations.

6 Is the ‘Fleeing’ Experience a System 1 Output?

Thus far, I have argued that the ‘fleeing’ experience is not a judgment because judgments are rationally revisable in response to evidence, whereas the ‘fleeing’ experience is not. This result borrows from the revisability view of judgment, on which all judgments must be capable of being rationally revised. Given that the revisability view of judgment excludes unrevisable cognitive states from the class of judgment, one might wonder whether one of these unrevisable non-doxastic cognitive attitudes might explain the ‘fleeing’ experience. For instance, one might wonder

whether System 1 outputs or delusion-like cognitive states might be unrevisable in a way that might make them suitable as an explanation of the ‘fleeing’ experience. In this section, I consider the prospects of explaining the ‘fleeing’ experience in terms of a System 1 output and argue that it cannot explain the ‘fleeing’ experience. In the next section, I consider a delusion-based explanation of the ‘fleeing’ experience and also find it lacking.

Recall that System 1 outputs are cognitive states which are produced quickly, heuristically, and automatically. They contrast with System 2 outputs, which are produced slowly, analytically, and effortfully. I have already argued that a version of the ‘fleeing’ experience which appeals to System 1 outputs is particularly attractive, as it enjoys the resources to account for the ‘fleeing’ experience’s automaticity and felt immediacy. Though I presumed in that section that System 1 outputs are a species of judgment, one might wonder whether the revisability view of judgment permits System 1 outputs into the class of judgment. The revisability view says that all beliefs must have a certain capacity to be rationally revised in response to conflicting evidence, and given the automatic and non-inferential *formation* of System 1 outputs, it might be thought that such outputs also lack the relevant capacity for subsequent revision.

In this section, I will present a dilemma for the view that the ‘fleeing’ experience is a System 1 output. Consistent with the empirical evidence, there are two plausible views of how System 1 outputs are revised in response to conflicting evidence: on the first view, they are rationally revised in response to conflicting evidence, in that they are either diminished in strength or else are extinguished altogether in response to such evidence. On the second view, System 1 outputs are never revised in response to conflicting evidence. Rather, when System 1 outputs conflict with evidence, System 1 outputs become non-conscious, but they are undiminished in strength.

For present purposes, we need not settle which of these empirically plausible accounts of how System 1 outputs are altered in response to evidence. On either of these accounts, the ‘fleeing’

experience is not a System 1 output. If on the one hand, System 1 outputs are revisable in response to the evidence, then System 1 outputs cannot explain the ‘fleeing’ experience for the same reason that judgments more generally cannot explain the ‘fleeing’ experience: System 1 outputs are capable of being revised, and the ‘fleeing’ experience is not. If on the other hand, System 1 outputs are not revisable, but are rather moved into the non-conscious in the face of conflicting evidence, it is also the case that System 1 outputs are unsuited to explain the ‘fleeing’ experience. Since the ‘fleeing’ experience remains conscious at the very same time as the judgment that conflicts with it, it does not become non-conscious in response to conflicting evidence and hence, is unlikely to be a System 1 output.

6.1 The First Horn of the Dilemma: System 1 Outputs as Rationally Revisable

There is some evidence to suggest that System 1 outputs are rationally revisable in response to conflicting evidence. I adduce two pieces of evidence that support this view: the first is evidence that typical System 1 outputs are governed by a cluster of mechanisms whose likely role is the detection of System 1/System 2 conflicts and the subsequent resolution of such conflicts. The second is a careful study of subject’s own reports in the Monty Hall Problem, which is a classic instance of a System 1/System 2 conflict problem. These reports suggest that subjects are capable of revising the tempting but erroneous System 1 output that this problem elicits.

There are some recent developments in the metacognitive literature which suggest the presence of a cluster of mechanisms whose likely function is to root out System 1/System 2 conflicts. I take the existence of these mechanisms to suggest some reason to think that System 1 outputs can be revised. The evidence for these mechanisms comes from the emerging literature on the metacognition of dual-systems reasoning and in particular from a recent result due to Valerie Thompson and Stephen Johnson.

In the Thompson and Johnson study, subjects were presented with a series of cognitive problems of the kind that tend to elicit System 1-produced responses. In some of these problems the correct answer to the problem and the intuitive, System 1-produced response were in conflict (*conflict problems*). In other problems, the correct answer to the problem and the intuitive, System 1-produced response were not in conflict (*no-conflict problems*). For each presented problem, subjects were asked to read the problem and then to indicate the first answer that came to mind. After doing so, they were asked to indicate by a 7-point numerical scale how strongly they felt the answer they had given was correct. The researchers took this response to correspond to subjects' *feeling of rightness*. They were then permitted to take as much further time as they liked to produce a second and final answer to the problem (Thompson & Johnson, 2014).

What the researchers found is that subjects reported a stronger feeling of rightness on no-conflict problems than on conflict problems, regardless of whether the answer they initially provided was in fact correct. Further, the weak feelings of rightness predicted a longer time spent on determining a second answer and a greater probability of a changed second answer, both of which are markers of analytic, System 2 reasoning. Strong feelings of rightness predicted a shorter time spent before making a second assessment and a reduced probability of changing one's answer (Thompson & Johnson, 2014). The so-called feeling of rightness turns out, then, *not* to track whether one answer is in fact correct but whether there is a conflict between what System 1 'suggests' and what System 1 'suggests.' This signaling of such a conflict in turn catalyzes increased analytic thought.

The Thompson and Johnson study suggests the presence of a cluster of related mechanisms that together plausibly ground our capacity for revising System 1 judgments. This is the mechanism of a conflict detector, which in turn triggers an affective response, which in turn catalyzes increased analytic engagement. Not only does the presence of such a mechanism make

plausible how System 1 outputs might be rationally revisable, if System 1 outputs were not rationally revisable, it would be difficult to explain why this cluster of mechanisms should exist: if not to root out conflicts, it is hard to see why conflict detection should exist and especially why it should trigger a particular feeling. And if not to *resolve* the detected conflict, it is hard to see why analytical engagement should be triggered by that feeling.⁴⁷ The presence of such a mechanism does not conceptually entail that System 1 outputs are capable of being revised, but it provides some reason to think they are.

The second bit of evidence that System 1 outputs are revisable comes from subjects' own reports about the Monty Hall problem. I mentioned previously that judgments can conflict and offered as an example a hypothetical subject who simultaneously enjoys the intuitive but incorrect judgment about the Monty Hall Problem, at the very same time that she judges that this solution is incorrect. Recall that the Monty Hall Problem is the problem of whether in a certain game show scenario involving hidden prizes and closed doors, you should switch from a door you have chosen or whether it does not matter whether you switch (for the details revisit section 4). It turns out that you should switch doors. This result is extremely counter-intuitive, as demonstrated by the fact that when the writer Marilyn vos Savant presented the problem along with its correct solution in a popular column, hundreds of readers wrote in to criticize her position (Rosenhouse, 2009). As it turns out, the incorrect, intuitive solution to the Monty Hall Problem is likely a System 1

⁴⁷ This resolution of the conflict may not always be in the direction of increased accuracy. Indeed, in the Thompson and Johnson study, increased analytical thinking—as signaled by a greater response time before the second answer and an increased probability of changing answers—did not increase the chance of a correct second answer. As Thompson and Johnson noted, this might be explained by the difficulty of many of the problems used in their study: ‘...the probability of normative responses overall was often observed to be at or near chance levels’ (Thompson and Johnson, 2014, p. 217).

output. In particular, it appears to be produced by an *equiprobability heuristic* (Saenen, Heyvaert, Van Dooren, & Onghena, 2015).

Interestingly, after vos Savant wrote several columns on the problem, most readers retracted their views. By the end of the Monty Hall debacle, about half of the letters to vos Savant were from readers who explicitly identified themselves as having previously endorsed the incorrect solution but as now endorsing the correct solution (Rosenhouse, 2009).

It is plausible that some of those who wrote to vos Savant to say they had changed their minds went through some period during which they simultaneously endorsed both the incorrect solution to the problem and the correct solution, and each to a high degree. But some of the letter authors appear to have moved past this (hypothesized) stage of cognitive dissonance, into a stage during which they endorse only the System-2 produced assessment, ‘you should switch doors’ and reject the System-1 produced assessment ‘it doesn’t matter whether you switch doors.’ Or at least, if these authors continue to endorse the System-1 assessment, they appear to hold it only with a very low degree of confidence. Consider two such letters:

You are indeed correct. My colleagues at work had a ball with this problem, and I dare say that most of them, including me at first, thought you were wrong!

The teachers in my graduate-level mathematics classes, most of whom thought you were wrong, conducted your experiment as a class project. Each of the twenty-five teachers had students in their middle or high school classes play at least 400 games. In all, we had 14,800 samples of the experiment, and we’re convinced that you were correct — the contestant should switch!⁴⁸

Notice that these authors describe their initial response as in the past tense (‘I... thought you were wrong’) and describe themselves as presently ‘convinced’ of the correct solution, which

⁴⁸ <http://marilynvosavant.com/game-show-problem/> Accessed March 14, 2015.

suggests a high degree of confidence. They also enthusiastically endorse the correct solution, which also suggests a high degree of confidence ('the contestant should switch!'). What is lacking from these authors' descriptions is any indication that they continue to hold—or at least continue to hold with much confidence—the counter-intuitive solution. Thus, if these authors' own reports are to be taken seriously, it seems plausible that they have revised or altogether relinquished their erstwhile System-1 produced assessment, suggesting its revisability. This is not to suggest that these letter writers might never again be tempted by the false solution. It is not as though they have somehow banished from their cognitive architecture the very system that grounds the tendency to respond to the Monty Hall Problem in the intuitive but incorrect way. They might in the future forget all about the Monty Hall Problem and encounter a variant of it elsewhere and again be tempted by or even endorse the incorrect solution. But at least *when they wrote these letters*, they appear to reject the intuitive System 1 assessment. This suggests that even this highly tempting and paradigmatic System-1 response is revisable in response to evidence.

In sum, the view that System 1 outputs are rationally revisable is at least defeasibly supported both by evidence of a cluster of mechanisms whose presumable function is the revision of System 1 outputs, and by subjects' own reports about the Monty Hall Problem. If System 1 outputs are revisable in this way, it is extremely implausible that the 'fleeing' experience should be a System 1 output. The 'fleeing' experience is characteristically unresponsive to evidence that conflicts with it, so if System 1 outputs are revisable in response to evidence, they are unsuited as an explanation of the 'fleeing' experience, for the same reason that judgments more generally are.

6.2 The Second Horn of the Dilemma: System 1 Outputs as Non-consciously Persisting

There is a second empirically plausible view of the relation between System 1 outputs and conflicting evidence, on which System 1 outputs are never strictly revised in response to such evidence, but are merely moved into the non-conscious in response to such evidence. On this view,

evidentially challenged System 1 outputs retain their strength but, if conscious, become non-conscious. This view is compatible with subjects' reports about the Monty Hall problem. On this view, the reason subjects convinced of the correct solution to the problem no longer take themselves to endorse the incorrect, System 1 assessment is that this assessment has become non-conscious and hence is not accessed by these subjects. This view is also compatible with the evidence of a cluster mechanisms designed to root out System 1/System 2 conflicts. On this view, the function of this cluster of mechanisms is to resolve (or perhaps, 'resolve') the conflict by moving System 1 outputs into the non-conscious, where they will ultimately degrade for reasons of processing capacity but not for reasons that are directly related to evidence.

As it turns out, there is also positive evidence to indicate that subjects who face cognitive tasks that trigger incongruent System 1/System 2 responses—tasks such as the Monty Hall Problem—and who correctly solve these tasks in favor of the System 2 response continue to token the rejected System 1 response at a non-conscious level. The evidence for this comes from subjects' performances on lexical decision tasks and other priming tasks designed to tease out these subjects' non-conscious representations; on such tasks, some of these subjects evince responses that are congruent with the disavowed System 1 response (Svedholm-Häkkinen, 2015).

Let's suppose that in fact, System 1 outputs are not revised in response to evidence but are at most pushed into the non-conscious by the advent of an incongruent System 2 output. This result would not make System 1 outputs a plausible candidate for explaining the 'fleeing' experience. For what is distinctive about the 'fleeing' experience—what explains why people find it so amusing and so surprising—is that the 'fleeing' interpretation is consciously accessed at the very same time as the incongruent judgment 'the smaller shapes aren't fleeing the triangle.' If the 'fleeing' experience were a System 1 output, then on this view of how System 1 outputs are altered in response to evidence, we should expect the 'fleeing' experience to become non-conscious. But the

‘fleeing’ experience remains conscious despite the evidence that it is inaccurate. Hence, it is extremely unlikely that the ‘fleeing’ experience is a System 1 output.

To summarize this section’s points: the empirical evidence suggests that when contravened by conflicting evidence, System 1 outputs either are revised or else are moved to the non-conscious, where they ultimately degrade for reasons of capacity limitations. If on the one hand, System 1 outputs are revisable—if they can be extinguished or reduced in strength in response to conflicting evidence—then the ‘fleeing’ experience cannot be a System 1 output since the ‘fleeing’ experience is not extinguished or diminished in response to conflicting evidence. If on the other hand, System 1 outputs can at most be pushed into the non-conscious in response to conflicting evidence but never truly extinguished, this too is a reason to think the ‘fleeing’ experience is not a System 1 output. Since the ‘fleeing’ experience is consciously accessed at the very same time that it is contravened by excellent evidence, it is not pushed into the non-conscious by such evidence and hence, is not a System 1 output.

7 Is The ‘Fleeing’ Experience a Delusion-like Cognitive Attitude?

In this section, I describe a different cognitive explanation of the ‘fleeing’ experience, *the delusion-based explanation*. On this view, the ‘fleeing’ experience is a delusion-like cognitive attitude and not a visual experience. I criticize this view on the grounds that the mechanisms that explain why delusions are not revisable cannot explain why the ‘fleeing’ experience is not revisable.

On the delusion-based explanation of the ‘fleeing’ experience, the experience is a delusion-like cognitive attitude with the content ‘the smaller shapes are fleeing the triangle.’ This attitude is not itself a visual experience, but is automatically and quickly generated in response to basic visual information, such as shape, color, and movement information. On this view, the ‘fleeing’ experience is not a species of delusion but is relevantly similar to delusions (this view is espoused

by Alex Byrne to account for the full range of visual illusions: Byrne, 2012, pp. 106-7; also see Byrne, 2011, p. 120).

It is important to clarify a weaker and stronger sense of the view that the ‘fleeing’ experience is a delusion-like cognitive attitude. On a weak version of this view, the ‘fleeing’ experience and the typical delusion both represent the world as being a certain way and both are unrevisable in response to evidence. I take this view to be true, and it is not my target. On a strong version of this view, the ‘fleeing’ experience is a delusion-like cognitive attitude, *and this is the most explanatorily fundamental category to which the ‘fleeing’ experience belongs*. I take this strong version of the view to be my target. As against this view, I will argue that the mechanisms that underlie the unrevisability of delusions are not operative in the typical subject enjoying the ‘fleeing’ experience and hence, if the most explanatorily fundamental thing we can say to explain why the ‘fleeing’ experience is unrevisable is that it is delusion-like, we are forced to accept as our best explanation a very explanatorily limited explanation. Ultimately, I will argue that we are not forced to this explanation, since we can instead adopt the more explanatorily fruitful view that the ‘fleeing’ experience is a visual illusion.

Though the correct characterization of delusions is controversial, the view of delusions that makes a delusion-based explanation of the ‘fleeing’ experience most plausible is one on which delusions are cognitive attitudes that persist despite what would appear to be excellent evidence against them. Delusions frequently involve contents that seem bizarre to the outside observer, for instance: that one does not exist or has died (*Cotard’s syndrome*), that certain of one’s thoughts are not one’s own (*thought insertion*), or that certain of one’s loved ones have been replaced by an imposter (*Capgras syndrome*). Though some delusions are the result of psychosis or other unusual perceptual phenomena, some delusions do not appear to be mediated by aberrant perception (Bell, Halligan, & Ellis, 2006, pp. 220-1).

7.1 Similarities between Delusions and the 'Fleeing' Experience

There are at least three respects in which delusions are on the face of it an apt model for the 'fleeing' experience: first, like the 'fleeing' experience, the typical delusion is plausibly a cognitive attitude, in that it represents the world as being a certain way. Second, both the 'fleeing' experience and the typical delusion are generated automatically, in response to stimuli, and as the result of a non-conscious transition. Finally—and, for our purposes, most importantly—both the 'fleeing' experience and the typical delusion are unresponsive to evidence that conflicts with them.

We can distinguish between *cognitive attitudes* and *conative attitudes* on the grounds that cognitive attitudes treat their objects as *obtaining*, whereas conative attitudes treat their objects as *to be obtained*. Judgments and beliefs are cognitive attitudes, as are merely entertained thoughts and cognitive imaginings. Wishes and hopes are paradigmatic conative attitudes (Shah & Velleman, 2005, p. 497).

Clinical delusions appear to be cognitive attitudes, i.e., to represent the world as a certain way. Delusional patients explicitly endorse the contents of their delusions and—in at least some cases—are guided in their actions by their delusions (So, et al., 2012, pp. 131-3; Wessely, 1993). Both of these attributes are easily explained if the delusion treats some state of affairs as obtaining, but is more difficult to explain if delusions merely represent some state of affairs as to be obtained.

A second commonality between the 'fleeing' experience and the typical delusion is that both are formed: in response to a stimulus, automatically, and unmediated by a conscious transition. We have already noted that it is a distinctive aspect of the 'fleeing' experience that the experience is formed in response to more basic visual information, automatically, and via a non-conscious transition. Likewise, delusions are frequently (and perhaps always) formed in response to

anomalous experiences, which are hallucinations or other unusual sensory experiences. Delusions are also frequently taken to arise automatically, i.e., without any initiation by their subject, and unmediated by a conscious transition (Fine, Craigie, & Gold, 2005).

A final similarity between the ‘fleeing’ experience and the typical delusion is that both are unresponsive to what would appear to be excellent evidence that conflicts with them. This aspect of delusions is on display in the following exchange between an examiner and a patient suffering from the delusion that his left hand is not his own:

Examiner: Whose hand is this?

Patient: Your hand.

(The examiner then places the patient’s hand between his own two hands)

Examiner: Whose hands are these?

Patient: Your hands.

Examiner: How many of them?

Patient: Three.

Examiner: Ever seen a man with three hands?

Patient: A hand is the extremity of an arm. Since you have three arms it follows that you must have three hands (Bisiach, 1998).

What is so fascinating about this dialogue is that the patient seamlessly invokes a valid inference pattern (‘Since you have three arms it follows that you must have three hands’) at the very same time that he casually relies on a premise that is both bizarre in character and patently false (‘you have three arms’). In this respect, this dialogue is an exemplar of what we know about delusional thought patterns more generally: while delusionals tend to exhibit roughly typical *inferential* capacities, they treat as indubitable claims that most would regard to be obviously false.

In the next section, I will argue that, despite the superficial similarities between the typical delusion and the ‘fleeing’ experience, the mechanisms that explain why delusions are not responsive to evidence are not operative in typical subjects and hence, cannot help explain why the typical subject’s ‘fleeing’ experience is not revised in response to evidence.

7.2 *Against the Delusion-based Explanation*

As many theorists have by now noted, a complete theory of delusions must explain both why delusions occur in the first place and why delusions persist despite what would appear to be excellent evidence against them (Cotlheart, 2007, p. 1044). We know that many delusions are *triggered* by anomalous experiences, but there are two reasons to think that anomalous experiences are not sufficient to explain the persistence of delusions: first, some *non*-delusional subjects have anomalous experiences. This raises the possibility that at least some of these subjects suffered an associated delusion at the onset of their anomalous experience, but that this delusion for some reason failed to persist.⁴⁹ Second, at least some delusional patients lack associated anomalous experiences. Either these patients enjoyed an anomalous experience in the past that triggered their delusion and their delusion outlived this experience, or else these patients' delusions were neither triggered nor sustained by anomalous experiences. In either case, anomalous experiences cannot explain the persistence of these patients' delusions (Bell, Halligan, & Ellis, 2008).

In light of the failure of anomalous experiences to account for delusional persistence, we require some other explanation. In the recent literature, the two most promising hypotheses about the persistence of delusions concern certain unusual reasoning biases and—in certain kinds of delusions—high levels of anxiety.

Although delusionals sometimes propose odd explanations to account for evidence, delusionals nevertheless exhibit roughly typical inferential capacities. Across a number of inductive and deductive reasoning tasks, delusionals perform indistinguishably from, or very similarly to,

⁴⁹ The other possibility is that none of these subjects' anomalous experiences triggered a delusion in the first place, which would suggest that anomalous experiences cannot wholly explain why delusions are triggered.

their non-delusional counterparts (Mujica-Parodi, Malaspina, & Sackeim, 2000). Despite this inferential typicality, delusionals exhibit certain other kinds of general reasoning biases that distinguish them from their non-delusional counterparts.⁵⁰ In particular, delusionals are more likely than their non-delusional counterparts to exhibit the traits of *jumping to conclusions* and *belief inflexibility*. These traits are general in that delusionals evince them even when reasoning about matters outside the scope of their delusions.

The trait of jumping to conclusions has been characterized as a tendency to gather less evidence than controls before forming an opinion. The evidence that delusional patients jump to conclusions comes from studies that compare how delusional and non-delusional patients perform tasks of probabilistic reasoning. In the classic ‘beads’ paradigm, subjects are shown two jars with opposite but equal ratios of beads: one contains 85 green beads and 15 pink beads, and the other contains 85 pink beads and 15 green beads. Subjects are also informed of the respective ratios of the beads. The jars are then concealed and patients are presented with beads drawn one at a time from one of the two jars. After each draw, subjects are free to make an assessment about which jar the bead was drawn from or to ask for another bead drawn before making an assessment. Delusional patients are more likely than non-delusional subjects to make an assessment after fewer draws, and in particular after *two or fewer* draws, which for the purposes of this paradigm is what is required to count as jumping to conclusions (Freeman, 2007, p. 437). In a meta-

⁵⁰ In calling these reasoning traits ‘biases,’ I am following the literature, but it is important to note that these traits are not necessarily epistemically or rationally inapt, nor is it necessary for the present argument that these traits be epistemically or rationally inapt. All that is required for my purposes is that delusionals evince *different* cognitive tendencies than non-delusionals, and that these tendencies contribute to the persistence of their delusions.

analysis of several ‘beads’ paradigms, it was found that between one-half and two-thirds of delusional subjects jump to conclusions, as compared to only 10% of non-delusional subjects (Freeman, 2007). Importantly, when delusional subjects are not given a choice about when to make an assessment, but are forced to consider a sufficiently large number of beads, delusional subjects’ performance on the probabilistic task is indistinguishable from that of typical subjects (Dudley, John, Young, & Over, 1997; cf. Garety & Freeman, 1999, p. 136). This suggests that delusional subjects’ capacity for probabilistic inference is typical, and that their different performance is due to their tendency to gather less evidence than controls before making an assessment.

Another general bias delusional subjects exhibit is *belief inflexibility*. In one study, belief inflexibility was operationalized as answering the following question negatively and with little hesitation, ‘when you think about it, is it at all possible that you might be mistaken in your belief?’ (So et al., 2012, p. 130). When asked about beliefs outside of the scope of their delusions, delusional subjects were more likely than non-delusional subjects to answer this question negatively and without hesitation (Colbert, Peters, & Garety, 2010).

There is emerging evidence that reasoning biases play at least a partial role in the maintenance of delusions. In three recent studies, delusional subjects underwent training designed to target general reasoning biases, including jumping to conclusions and belief inflexibility (Favrod et al., 2014; Garety et al., 2014; Moritz et al., 2013). In all three studies, subjects who underwent reasoning training exhibited a significant reduction in their reasoning biases and in delusional strength, as compared to a control group that did not receive the reasoning training. These results would be easily explained if general reasoning biases contributed to delusional persistence.⁵¹

⁵¹ Better evidence that general reasoning biases contribute to delusional persistence would be evidence that a reduction in reasoning biases mediates a reduction in delusionality. In the Garety-led study, there was some evidence that a reduction in reasoning biases mediated a reduction in

Another factor that contributes to the maintenance of certain delusions—and in particular of *persecutory delusions*—is high levels of worry. Persecutory delusions are delusions that others are trying to harm one and are the most common of all delusions (Cannon & Kramer, 2012, pp. 323-4). Levels of worry among those with persecutory delusions are similar to those with generalized anxiety disorder (Freeman & Garety, 2014, p. 1180). In a ground-breaking interventional study, Daniel Freeman and colleagues showed worry reduction therapy reduces the strength of persecutory delusions and further, that ‘changes in worry mediated the majority of the change in the delusions’ (Freeman & Garety, 2014, p. 1180, discussing Freeman et al., 2014). This result provides excellent evidence that the persistence of persecutory delusions are at least partly caused by high levels of worry, though the precise mechanism by which worry contributes to delusional persistence is not yet understood.

Since the best current explanations of why delusions are unrevisable proceed in terms of reasoning biases and atypically high levels of worry, delusions are ill-suited as a model for the ‘fleeing’ experience. The ‘fleeing’ experience is elicitable in almost all typical subjects. Thus, for at least many such subjects, that the ‘fleeing’ experience is not revised in response to the evidence that conflicts with it cannot be explained by reasoning biases or high levels of worry. We can be confident that at least many subjects without either reasoning biases or high levels of worry enjoy the ‘fleeing’ experience when viewing the Heider and Simmel movie.

It might be objected, quite reasonably, that there remains the possibility that certain delusions persist for reasons that are explained neither by general reasoning biases nor by high levels of worry. The question for the advocate of the delusion-based model of the ‘fleeing’ experience is why we should think that the mechanisms that explain why these (hypothesized) delusions are

delusional, but after controlling for confounds, this effect fell just below statistical significance (Garety et al., 2014).

not revisable should also apply to the typical subject undergoing the ‘fleeing’ experience. Should such delusions turn out to persist for reasons that are not relevant to the typical subject viewing the Heider and Simmel movie, they will be no more equipped than those whose mechanisms we partly understand to explain why this subjects’ ‘fleeing’ experience is unrevisable. In conclusion, to insist that the ‘fleeing’ experience is a delusion-like state, and that this is the most explanatorily fundamental mental kind to which it belongs, would be to leave us without any satisfying explanation of the ‘fleeing’ experience’s unrevisability.

In light of the failure of both System 1 outputs and delusions to help explain the ‘fleeing’ experience, the advocate of a cognitive approach to the ‘fleeing’ experience might maintain that the ‘fleeing’ experience is a *sui generis* cognitive attitude, one which is neither reducible to nor even helpfully modeled on other cognitive attitudes. This is a conceptual possibility that cannot be excluded by any of the broadly empirical considerations I’ve offered. If there were no other available explanations of the ‘fleeing’ experience, this interpretation might seem plausible, even inevitable. But, as I will suggest in the next section, the visual explanation of the ‘fleeing’ experience enjoys the resources to account for the unrevisability of the ‘fleeing’ experience. Thus, we are not forced to posit that the ‘fleeing’ experience is a novel cognitive attitude. If we conceive of the ‘fleeing’ experience as a visual experience, on a par with visual experiences of shape and color, its unrevisability is easily accounted for.

8 Visually Perceiving the Intentions of Others

Thus far, I have argued that the ‘fleeing’ experience is automatic, stimulus-driven, and dependent on more basic visual information such that it is least *prima facie* plausible that the experience is a visual experience. I have further argued that this experience is neither a post-perceptual judgment nor a System 1 judgment nor a delusion-like cognitive state.

Now I would like to suggest that the visual explanation of the ‘fleeing’ experience succeeds where the considered cognitive explanations failed. On the visual explanation, the ‘fleeing’ experience is a visual experience. To be more precise, the ‘fleeing’ experience is a visual illusion, since it misrepresents some aspect of the world. If we claim that the ‘fleeing’ experience is a visual illusion, we can assimilate it into the class of already established visual experiences that do not respond to at least some kinds of relevant conflicting evidence, such as Titchener’s Illusion, the Müller-Lyer Illusion, or Aristotle’s ‘bent’ stick. If the ‘fleeing’ experience just is a kind of visual illusion, we are not forced to posit a *sui generis* cognitive attitude to explain it. Nor are we forced to accept that the most explanatorily fundamental category to which the ‘fleeing’ experience belongs is that of *delusion-like cognitive attitude*.

The mechanisms that explain why visual illusions in general are not rationally responsive to conflicting evidence also plausibly govern the typical subject’s ‘fleeing’ experience and hence, the visual explanation can provide a satisfying account of the ‘fleeing’ experience. There are two major accounts of why visual illusions are not rationally responsive to at least some kinds of evidence. On one view, some kinds of extra-visual information never enter into the visual system as inputs, and this explains why visual experiences are unrevised in response to these sorts of information. On this view, the visual system enjoys the capacity to process extra-visual information, but the visual system lacks *access* to extra-visual representations, which prevents its processing them (Fodor, 1983; Fodor, 1984; Fodor, 2001, p. 63). On another view, the visual system’s ability to process information is restricted in the kinds of information it can process, and it is this processing limitation which explains why the system cannot process certain kinds of extra-visual information. On this view, the visual system enjoys access to the relevant extra-visual representa-

tions, but it lacks the ability to process those representations, and this explains why visual experiences are rationally immune to these kinds of information (Barrett, 2006; Carruthers, 2004; Carruthers, 2006; Carruthers, 2011).

Thus, there are two plausible explanations of why visual illusions are rationally immune to certain kinds of evidence, one of which appeals to a lack of access to the relevant evidence and one of which appeals to a processing limitation of the visual system itself. For our purposes, it does not matter which of these explanations is correct. Since both appeal to mechanisms supposed to be present in typical subjects, both are apt as explanations of the typical subject's 'fleeing' experience. On the first view, the reason the 'fleeing' experience is unresponsive to the simultaneously held judgment that it is inaccurate is that the 'fleeing' experience is a visual experience, and the visual system lacks access to this judgment. On the second view, the reason the 'fleeing' experience is unresponsive to the simultaneously held conflicting judgment is that the visual system lacks the capacity to process this judgment. In either case, if the 'fleeing' experience is a visual experience, we can explain its unrevisability by appealing to the mechanisms that underlie visual illusions more generally.

BIBLIOGRAPHY

- Adler, J. (2002). *Belief's own ethics*. Cambridge, MA: MIT Press.
- Audi, R. (2013). *Moral perception*. Princeton: Princeton University Press.
- Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, 26(4), 313-331.
- Bayne, T. (2008). The phenomenology of agency. *Philosophy Compass*, 3(1), 182-202.
- Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly*, 59(236), 385-404.
- Bayne, T. (2010). *The unity of consciousness*. Oxford: Oxford University Press.
- Bayne, T. (2011). The sense of agency. In F. Macpherson (Ed.), *The senses* (355-374). New York: Oxford University Press.
- Bayne, T. & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind & Language*, 20(2): 163-188.
- Beier, J. S. & Carey, S. (2014). Contingency is not enough: Social context guides third-party attributions of intentional agency. *Developmental Psychology*, 50(3), 889-902.
- Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Explaining delusions: a cognitive perspective. *Trends in Cognitive Sciences*, 10(5), 219-226.
- Bell, V., Halligan, P. W., & Ellis, H. D. (2008). Are anomalous perceptual experiences necessary for delusions? *The Journal of Nervous and Mental Disease*, 196(1), 3-8.
- Bird, A. (1998). Dispositions and antidotes. *The Philosophical Quarterly* 48, 227-234.
- Block, N. (1996). Mental paint and mental latex. *Philosophical Issues* 7, 19-49.

- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 30, 481-499.
- Block, N. (2010). Attention and mental paint. *Philosophical Issues*, 20(1), 23-63.
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567-575.
- Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, 89(1): 650-572.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Brogaard, B. (2012). Do we perceive natural kind properties? *Philosophical Studies*, 1-8.
- Burge, T. (2010). Origins of perception. *Disputatio IV* (29), 1-38.
- Butterfill, S. (2009). Seeing causings and hearing gestures. *The Philosophical Quarterly*, 59(236), 405-428.
- Bykvist, K. & Hattiangadi, A. (2007). Does thought imply uught? *Analysis* 67, 277-285.
- Byrne, A. (2009). Experience and content. *The Philosophical Quarterly*, 59(236), 429-451.
- Byrne, A. (2011). Knowing that I am thinking. A. Hatzimoysis (ed.) *Self-knowledge*, (pp. 105-124). Oxford: Oxford University Press.
- Byrne, A. (2012). Knowing what I see. In D. Smithies & D. Stoljar (eds.), *Introspection and Consciousness*, Oxford: Oxford University Press.
- Cacioppo, J. T. & Petty, R. E. (1989). Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, 10(1), 3-12.
- Cannon, B. J., & Kramer, L. M. (2012). Delusion content across the 20th century in an American psychiatric hospital. *International Journal of Social Psychiatry*, 58(3), 323-327.
- Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
- Cassam, Q. (2010). Judging, believing, and thinking. *Philosophical Issues* (20)1, 80-95.

- Chalmers, D. (2006). Perception and the fall from Eden. In T. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (49-125). Oxford: Oxford University Press.
- Chappell, T. (2008). Moral perception. *Philosophy*, 83(04), 421-437.
- Chrisman, M. (2008). Ought to believe. *Journal of Philosophy* 105, 346 – 370.
- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3), 181-204.
- Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.
- Colbert, S. M., Peters, E. R., & Garety, P. A. (2010). Delusions and belief flexibility in psychosis. *Psychology and Psychotherapy: Theory, Research and Practice*, 83(1), 45-57.
- Coltheart, M. (2007). The 33rd Sir Frederick Bartlett lecture cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology*, 60(8), 1041-1062.
- Congiu, S., Schlottmann, A., & Ray, E. (2010). Unimpaired perception of social and physical causality, but impaired perception of animacy in high functioning children with autism. *Journal of autism and developmental disorders*, 40(1), 39-53.
- Crutchfield, P. (2012). Representing high-level properties in perceptual experience. *Philosophical Psychology*, 25(2), 279-294.
- Cullison, A. (2010). Moral perception. *European Journal of Philosophy*, 18(2), 159-175.
- Currie, G. and Ravenscroft, I. (2002): *Recreative minds*. Oxford: Oxford University Press.
- Davidson, D. (1984). *Inquiries into truth and interpretation*. Clarendon: Oxford University Press.
- Davies, M. & Coltheart, M. (2000). Introduction: pathologies of belief. *Mind & Language* 15(1), 1-46.
- Dennett, D. C. (1989). *The intentional stance*. Cambridge, MA: MIT Press.
- Doggett, T., & Egan, A. (2007). Wanting things you don't want. *Philosophers' Imprint* 7(9), 1-17.
- Döring, F. (1990). Limits of irrationality. *Philosophy and Psychopathology*, 86-101.

- Dretske, F. I. (1979). Simple seeing. In *Body, mind, and method* (1-15). Springer Netherlands.
- Dudley, R. E. J., John, C. H., Young, A. W., & Over, D. E. (1997). Normal and abnormal reasoning in people with delusions. *British Journal of Clinical Psychology*, 36(2), 243-258.
- Egan, A. (2009). Imagination, delusion, and self-deception. in *Delusion and Self-deception: Affective and Motivational Influences on Belief Formation*, eds. T. Bayne and J. Fernandez. New York: Psychology Press.
- Elliot, A. & Devine, P. (1994). On the motivational nature of cognitive dissonance: dissonance as psychological discomfort. *Journal of Personality and Social Psychology* 67, 382-394.
- Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59, 255-278.
- Farennikova, A. (2013). Seeing absence. *Philosophical studies*, 166(3), 429-454.
- Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press
- Favrod, J., Rexhaj, S., Bardy, S., Ferrari, P., Hayoz, C., Moritz, S., & Bonsack, C. (2014). Sustained antipsychotic effect of metacognitive training in psychosis: A randomized-controlled study. *European Psychiatry*, 29(5), 275-281.
- Feldman, R. (2000). The ethics of belief. *Philosophy and Phenomenological Research* 60, 667-695.
- Fine, C., Craigie, J., & Gold, I. (2005). Damned if you do; damned if you don't: The impasse in cognitive accounts of the Capgras Delusion. *Philosophy, Psychiatry, & Psychology*, 12(2), 143-151.
- Fish, W. (2012). High-level properties and visual experience. *Philosophical Studies*, 1-13.

- Freeman, D. (2007). Suspicious minds: the psychology of persecutory delusions. *Clinical Psychology Review*, 27(4), 425-457.
- Freeman, D., Dunn G., Startup, H., Pugh, K., Cordwell, J., Mander, H., Cernis, E., Wingham, G., Shirvell, K., & Kingdon, D. (2014). An explanatory randomised controlled trial testing the effects of cognitive behaviour therapy for worry on persecutory delusions in psychosis: the Worry Intervention Trial (WIT) (submitted)
- Freeman, D. & Garety, P. (2014). Advances in understanding and treating Persecutory Delusions: A Review. *Soc Psychiatry Epidemiol.* 49(8), 1179–1189.
- Friedman, J. (2013). Suspended judgment. *Philosophical Studies* 162(2), 165-181.
- Frisby, J. P., & Stone, J. V. (2010). *Seeing: The computational approach to biological vision*. The MIT Press.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535-543.
- Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. *British journal of clinical psychology*, 38(2), 113-154.
- Garety, P. A., Gittins, M., Jolley, S., Bebbington, P., Dunn, G., Kuipers, E., & Freeman, D. (2013). Differences in cognitive and emotional processes between persecutory and grandiose delusions. *Schizophrenia Bulletin*, 39(3), 629-639.
- Garety, P. A., Hemsley, D. R., & Wessely, S. (1991). Reasoning in deluded schizophrenic and paranoid patients: biases in performance on a probabilistic inference task. *The Journal of Nervous and Mental Disease*, 179(4), 194-201.
- Garety, P., Waller, H., Emsley, R., Jolley, S., Kuipers, E., Bebbington, P. & Freeman, D. (2014). Cognitive mechanisms of change in delusions: an experimental investigation targeting reasoning to effect change in paranoia. *Schizophrenia Bulletin*.

- Gendler, T. S. (2007). Self-deception as pretense. *Philosophical Perspectives* 21(1), 231-258.
- Gendler, T. S. (2008a). Alief and belief. *The Journal of Philosophy*, 634-663.
- Gendler, T. S. (2008b). Alief in action (and reaction). *Mind & Language*, 23(5), 552-585.
- Gertler, B. (2011). Self-knowledge and the transparency of belief. *Self-Knowledge*. Oxford: Oxford.
- Glüer, K. (2009). In defence of a doxastic account of experience. *Mind & Language*, 24(3), 297-327.
- Glüer, K. & Wikforss, Å. (2013). Aiming at truth: on the role of belief. *Teorema*, 32(3), 137-162.
- Graham, P. A. (2011). 'Ought' and ability. *The Philosophical Review*, 120, 337- 382.
- Harmon-Jones, E. (2000). A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes. *Emotions and Beliefs*, 185-211.
- Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1), 7-16.
- Hashimoto, H. (1966). A phenomenal analysis of social perception. *Journal of Child Development*, 2, 1-26.
- Hattiangadi, A. (2010). The love of truth. *Studies in History and Philosophy of Science* 41, 422-432.
- Hazlett, A. (2013). *A luxury of the understanding: on the value of true belief*. Oxford: Oxford University Press.
- Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), 7487-7491.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 243-259.

- Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly* 87(1), 45-74.
- Hoerl, Christoph. (2001). On thought insertion. *Philosophy, Psychiatry, & Psychology* 8 (2-3), 189-200.
- Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
- Horgan, T., Tienson, J., & Graham, G. (2003). The phenomenology of first person agency, In *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, ed. Exeter, UK: Imprint Academic.
- Huddleston, A. (2012). Naughty beliefs. *Philosophical Studies*, 160(2), 209-222.
- Johnston, M. (1992). How to speak of the colors. *Philosophical Studies* 68, 221–263.
- Kelly, T. (2008). Evidence: fundamental concepts and the phenomenal conception. *Philosophy Compass* 3(5), 933-955.
- Koldewyn, K., Hanus, P., & Balas, B. (2013). Visual adaptation of the perception of “life:” Animacy is a basic perceptual dimension of faces. *Psychonomic Bulletin & Review*, 1-7.
- Lehrer, K. (1976). ‘Can’ in theory and practice: A possible world analysis.’ in *Action Theory*, eds. Brand and Walton. Dordrecht: D. Reidel.
- Leon, M. (1992). Rationalising belief. *Philosophical Papers*, 21(3), 299-314.
- Lewis, David. (1980). Veridical hallucination and prosthetic vision. *Australasian Journal of Philosophy* 58 (3), 239-249.
- Littlejohn, C. (2012). Does ‘ought’ still imply ‘can’? *Philosophia* 40(4), 821-828.
- Logue, H. (2012). Visual experience of natural kind properties: is there any fact of the matter? *Philosophical Studies*, 1-12.
- Macpherson, F. (2011). The admissible contents of experience. In F. Macpherson & K. Hawley (Eds.), *The Admissible Contents of Mental States*, Wiley-Blackwell.

- Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24-62.
- Mandelbaum, E. (2013). Against alief. *Philosophical Studies*, 165(1), 197-211.
- Mandelbaum, E. (2014). Thinking is believing. *Inquiry* 57(1), 55-96.
- Martin, J. R., & Dokic, J. (2013). Seeing absence or absence of seeing? *Thought: A Journal of Philosophy*, 2(2), 117-125.
- Masrour, F. (2011). Is perceptual phenomenology thin? *Philosophy and Phenomenological Research*, 83(2), 366-397.
- McHugh, C. (2011). What do we aim at when we believe? *Dialectica*, 65(3), 369-392.
- Mele, A. R., & Moser, P. K. (1994). Intentional action. *Noûs*, 39-68.
- Mizrahi, M. (2012). Does 'ought' imply 'can' from an epistemic point of view? *Philosophia* 40, 829-840.
- Moritz, S., Veckenstedt, R., Bohn, F., Hottenrott, B., Scheu, F., Randjbar, S., & Roesch-Ely, D. (2013). Complementary group Metacognitive Training (MCT) reduces delusional ideation in schizophrenia. *Schizophrenia Research*, 151(1), 61-69.
- Mujica-Parodi, L. R., Malaspina, D., & Sackeim, H. A. (2000). Logical processing, affect, and delusional thought in schizophrenia. *Harvard Review of Psychiatry*, 8(2), 73-83.
- Nanay, B. (2011). Do we see apples as edible? *Pacific Philosophical Quarterly* 92 (3), 305-322.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2(2), 175-220.
- Pautz, A. (2009). What are the contents of experiences? *The Philosophical Quarterly*, 59(236), 483-507.
- Phillips, Ian. forthcoming. Hearing and hallucinating silence. In F. Macpherson & D. Patchias (Eds.), *Hallucination* MIT Press.

- Pitt, David. (2004). The phenomenology of cognition; or *what is it like to think that p?* *Philosophy and Phenomenological Research* 69, 1–36.
- Prinz, J. (2013). Siegel’s get rich quick scheme. *Philosophical Studies*, 1-9.
- Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(03), 341-365.
- Rey, G. (2013). We are not all ‘self-blind’: A defense of a modest introspectionism. *Mind & Language*, 28(3), 259-285.
- Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, 6(1), 77-97.
- Rosenhouse, J. (2009). *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brain Teaser*. Oxford University Press.
- Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, 32(3), 189-194.
- Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies* 114(1), 47-79.
- Saenen, L., Heyvaert, M., Van Dooren, W., & Onghena, P. (2015). Inhibitory control in a notorious brain teaser: the Monty Hall dilemma. *ZDM*, 1-12.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Schlottmann, A. & Shanks, D.R. (1992). Evidence for a distinction between judged and perceived causality. *Quart. J. Exp. Psychol. A. Hum. Exp. Psychol.*, 44, 321–342.
- Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? *Social perception: Detection and interpretation of animacy, agency, and intention*. Cambridge, MA: MIT Press

- Scholl, B. J., & Tremoulet, P.D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299-309.
- Schwind, C., Buder, J., Cress, U., & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education* 58(2), 787-796.
- Schwitzgebel, E. (2001). In-between believing. *The Philosophical Quarterly* 51(202), 76-82.
- Schwitzgebel, E. (2002). A dispositional, phenomenal account of belief. *Noûs* 36(2), 249-275.
- Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review* 117 (2): 245-273.
- Schwitzgebel, E. (2012). Mad belief? *Neuroethics* 5(1), 13-17.
- Segall, M., Campbell, D., & Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*. New York: The Bobbs-Merrill Company.
- Setiya, K. (2014). Intention, *The Stanford Encyclopedia of Philosophy* (Spring 2014), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2014/entries/intention/>>.
- Shah, N., & Velleman, J. D. (2005). Doxastic deliberation. *The Philosophical Review*, 497-534.
- Shea, N. (2014). Distinguishing top-down from bottom-up effects, in *Perception and Its Modalities*. S. Biggs, M. Matthen, D. Stokes (eds.) Oxford: Oxford University Press.
- Siegel, S. (2005). The phenomenology of efficacy. *Philosophical Topics*, 33(1), 265-284.
- Siegel, S., (2006). Which properties are represented in perception? in *In Perceptual Experience* (4810-503). T. Gendler Szabo and J. Hawthorne (eds.), Oxford: Oxford University Press.
- Siegel, S. (2007). How can we discover the contents of experience? *The Southern Journal of Philosophy*, 45(S1), 127-142.
- Siegel, S. (2009). The visual experience of causation. *The Philosophical Quarterly*, 59(236), 519-540.

- Siegel, S. (2010). *The contents of visual experience*. Oxford: Oxford University Press.
- Siegel, Susanna, (2013). The contents of perception, *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/perception-contents/>>
- Silins, N. (2012). The significance of high-level content. *Philosophical Studies*, 1-21.
- Slaughter, V., Heron-Delaney, M., Christie, T., Slaughter, V., & Brownell, C. (2011). Developing expertise in human body perception. In Slaughter & Brownell (Eds.) *Early Development of Body Representations*, Cambridge: Cambridge University Press.
- So, S. H., Freeman, D., Dunn, G., Kapur, S., Kuipers, E., Bebbington, P., & Garety, P. A. (2012). Jumping to conclusions, a lack of belief flexibility and delusional conviction in psychosis: a longitudinal investigation of the structure, frequency, and relatedness of reasoning biases. *Journal of Abnormal Psychology*, 121(1), 129.
- Sorensen, R. (2012). The Twin Towers riddle. *Philosophical Studies*, 1-9.
- Spaulding, S. (forthcoming). On whether we can see intentions. *Pacific Philosophical Quarterly*.
- St. B. T. Evans. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59, 255-278.
- Steup, M. (2008). Doxastic freedom. *Synthese* 161(3), 375-392.
- Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 499-518.
- Stokes, D. (2014). Cognitive penetration and the perception of art. *Dialectica*, 68(1), 1-34.
- Strawson, G. (2011). Cognitive phenomenology: real life. In T. Bayne & M. Montague (Eds.) *Cognitive Phenomenology*, Oxford: Oxford University Press.
- Strevens, M. (2012). Notes on Bayesian confirmation theory. unpublished manuscript.
- Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta Psychologica*, 154, 69-76.

- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *In two minds: dual processes and beyond*. Oxford University Press, Oxford.
- Thompson, V. A. & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, 20(2).
- Tye, M. (2009). *Consciousness revisited: materialism without phenomenal concepts*. MIT Press.
- Van Gulick, R. (1994). Deficit studies and the function of phenomenal consciousness, in: G. Graham and G. L. Stephens (Eds.), *Philosophical Psychopathology*, 25-49. Cambridge, MA: MIT Press.
- van Leeuwen, D. N. (2009). The motivational role of belief. *Philosophical Papers* 38(2), 219-246.
- Velleman, D. (2000). *The possibility of practical reason*. Oxford: Oxford University Press.
- Vranas, P. (2007). I ought, therefore I can. *Philosophical Studies* 136, 167 - 216.
- Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., & Haudenschild, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behaviour Research and Therapy*, 45(6), 1255-1269.
- Way, J. (2010). The normativity of rationality. *Philosophy Compass* 5(12), 1057-1068.
- Wedgwood, R. (2002). The aim of belief. *Noûs*, 36(s16), 267-297.
- Wessely, S., Buchanan, A., Reed, A., Cutting, J., Everitt, B., Garety, P., & Taylor, P. J. (1993). Acting on delusions. I: Prevalence. *The British Journal of Psychiatry*, 163(1), 69-76.
- Whiting, D. (2010). Should I believe the truth? *Dialectica* 64: 213-224.
- Woodward, T. S., Munz, M., LeClerc, C., & Lecomte, T. (2009). Change in delusions is associated with change in “jumping to conclusions.” *Psychiatry Research*, 170(2), 124-127.