BACKGROUND INFLUENCES: PSYCHOLOGICAL PROCESSES THAT SHAPE PERCEPTION, EMOTION, AND MORALS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Jonathan Russell Vance
May 2013

BACKGROUND INFLUENCES: PSYCHOLOGICAL PROCESSES THAT SHAPE PERCEPTION, EMOTION, AND MORALS

Jonathan Russell Vance Ph. D.
Cornell University 2013

Abstract In this three chapter dissertation, I address the epistemic significance of psychological processes that shape emotion, perception, and moral judgment. In Chapter 1, I illustrate ways in which perceptual and emotional states can be influenced by background beliefs in a process called 'cognitive penetration'. I then use cases of cognitively penetrated emotion to provide a novel argument by analogy against views of perceptual justification (e.g. dogmatism) that emphasize the justificatory role of an experience's phenomenology rather than its etiology, including etiologies involving cognitive penetration. In Chapter 2, I extend the challenge from cognitive penetrability to target reliabilism, a view that emphasizes the justificatory role of etiology rather than phenomenology. In my view, both phenomenology and etiology have a role to play in fixing the justification provided by perceptual and emotional states. In Chapter 3, I turn to the covert influence of morally irrelevant factors on emotion and moral judgment. Recent authors have used empirical evidence of the influence of such factors to argue for the skeptical conclusion that intuitive moral judgments are unreliable. In response, I argue that the data indicate that the influences are too small to threaten the reliability of the relevant judgments, and in fact may provide novel support for an empirically plausible, moral epistemology that gives perception-like moral emotion a role in justifying intuitive moral beliefs.

BIOGRAPHICAL SKETCH

Jonathan (Jona) Vance grew up in Grand Rapids, Michigan. After graduating from Calvin College with a BA in Philosophy (Honors) in 2006, Jona spent a year working as an AmeriCorps Member at Albany Park Community Center in Chicago, Illinois. Jona is married to Rebecca Marie Rinsema. They have a son named Judah.

For Rebecca and Judah

ACKNOWLEDGMENTS

I have many people to thank for helping me produce this dissertation. I apologize to those whom I forget to thank but should have.

I thank my special committee members. Thanks to Derk Pereboom for one of my first and best graduate seminars, use of his spacious office while he was in Rome, books, and great advice. Thanks to Carl Ginet for pressing me clarify my positions on a numerous issues, comments in short order whenever I sent a draft, and many wonderful lunches at Banfi's. Thanks to David Pizarro—the best philosopher in a psychology department one will ever find, and a pretty good psychologist too—for wit, insight into the empirical literature, and great discussions on the way home after moral reasoning seminars. And to Nico Silins, my special committee chair: thanks for encouraging me from the beginning, for always being available to meet to discuss new work, and for giving so many detailed, incisive-but-generous comments on my work in writing and in person.

Thanks also to Andrew Chignell for including me in a project on hope and optimism, which helped me develop additional research projects beyond the dissertation and gave my family opportunities we would not have otherwise had.

Thanks to my fellow graduate students. Special thanks to Lu Teng and Ru Ye for good food and conversation, and, especially, to Adam Bendorf, Stephen Humphreys-Mahaffey, and Colin McLear for great conversations, powder days,

feedback on written work, and rides to the airport.

Thanks to Dorothy Vanderbilt and Paula Epps-Cepero for their patience and help on innumerable things.

Thanks to Caden Hare for his courage and for inviting me to hang out on his porch one summer to talk about Plato.

Thanks to the Cognitive Science Program at Cornell for a summer fellowship that supported work on the dissertation and to Cornell's Society for the Humanities for a dissertation writing grant. Thanks to Brie Gertler, Michelle Kosch, Jesse Prinz, and Nicholas Sturgeon for helpful feedback on earlier drafts of this material and to audiences at Cornell University, the University of Western Ontario, and the University of South Florida.

Thanks to Bob and Debbie Vance for loving and supporting me no matter what, and for much more. Thanks to Clyde and Beth Rinsema for their love and support as well. Thanks to my brother Matthew, who will be writing an acknowledgments section himself before too long, and to my sister Julia for having the strength to find a meaningful life outside academia.

Thank you to Judah for coming into my life. You are my favorite boy in the world. I will love and support you with everything I have.

Finally, to Rebecca: without you, I could not have done even a tiny fraction of what I have done. I love you, and I am so happy we get to have this life together.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation consists of three interrelated chapters on perception, emotion, and moral judgment. In the preface, I elaborate a number of the themes that tie the chapters together.

Each of the chapters in the dissertation concerns background influences by psychological states on subjects' experiences. The experiences include perceptual and emotional states. In Chapters 1 and 2, I illustrate ways in which perceptual and emotional states can be influenced by background beliefs in a process called 'cognitive penetration'. In Chapter 3, I illustrate ways in which moral attitudes can be influenced by emotions elicited from stimuli unrelated to the moral issues at stake. For example, moral attitudes can be affected by hypnotically induced disgust or anger caused by a film whose content is unrelated to the content of the moral attitude. Moral emotions may also be influenced by cognitive penetration.

The main theses of the dissertation concern the epistemic significance of these background influences. In most of the cases I discuss, subjects are unaware of the influence their background states have on their experiences. I argue that such states can have an impact on the justification and knowledge the relevant subjects' enjoy as the result of their experiences, even though the subjects are unaware of the causal processes that influence the experiences. The claim is controversial, and the dissertation makes novel contributions toward establishing it.

An important claim throughout the dissertation—but especially in Chapters 1 and 3—is the following: some emotional states are analogous to some perceptual states. I should note that I do not argue that all emotional states are analogous to perceptual states. Instead, I claim only that some pairs of such states are analogous. This weaker claim is sufficient for the arguments I make in the dissertation.

To develop the analogy between some emotional and perceptual states, I highlight and argue for the following claims. Some emotional and perceptual states have a distinctive kind of phenomenology, which I call 'presentational phenomenology'. Some of these emotional and perceptual states have similar etiologies; for example, both can result from cognitive penetration. Moreover, instances of both types of state can represent external world contents. Finally, both types of state can provide defeasible, epistemic justification under some circumstances. In some such circumstances the relevant justification is fixed in part by the states' phenomenology and in part by their etiology.

Arguing for the analogy between some emotional and perceptual states is philosophically important on its own. It also enables me to make novel arguments in the epistemology of perception and moral judgment. Regarding perception, in Chapter 1, I use the analogy to argue from relatively uncontroversial epistemic claims about emotions to controversial epistemic claims about perception. In Chapter 3, I use the analogy to argue from relatively uncontroversial epistemic claims about perception to controversial claims about moral judgment.

Throughout the dissertation, I take a naturalistic approach to philosophy. As I practice it, philosophy is continuous with the empirical sciences. The dissertation is especially engaged with the social and cognitive sciences. It draws extensively from and aims to contribute to work in both psychology and philosophy. In my view, results from empirical psychology can play an important role in philosophical theorizing.

I use empirical results in the dissertation in each of the following ways. In Chapter 1, I use data about the psychological processes underlying perception and emotion to argue that both perceptual and emotional states are cognitively penetrable. In turn, I use these data to support the analogy between some perceptual and emotional states. The data also provide evidence that the counterexamples to theories of epistemic justification discussed in the first two chapters are realistic. In Chapter 2, I use additional psychological data on the cognitive penetration of perception to argue that cognitively penetrated perceptual processing reliably produces true beliefs. In Chapter 3, I discuss data on the influence of emotion on moral judgment. I also discuss psychological methodology in collecting the data. I make observations about what the data do and do not show and offer recommendations for how best to conduct the relevant psychological studies to obtain philosophically useful information.

An additional theme regarding the use of empirical data emerges throughout the dissertation and is especially important in Chapters 2 and 3. It concerns the need

to interpret such data cautiously. For example, psychologists and philosophers alike have interpreted results concerning extraneously induced emotions as revealing frequent shifts in our moral judgments in response to morally irrelevant factors. However, I argue that a closer look at the data reveals that no such evidence is provided. The effect sizes of the relevant results are too small to provide evidence of frequent shifts in moral judgments. (I make similar points in Chapter 2 concerning empirical data on the reliability of judgments formed on the basis of cognitively penetrated perceptual experiences.) In addition, in Chapter 3, I note that there are important differences in the wording of survey question for various studies. As a result, survey questions intended to measure the influence of morally irrelevant factors on emotions and moral judgments may measure the influence on subtly but importantly different attitudes. For example, some questions seem to track subjects' degrees of confidence whereas others track their outright judgments. I note that these differences are significant for our understanding of moral cognition and for the philosophical arguments that utilize the relevant studies.

Although I make considerable use of empirical data and aim to contribute to psychological theory and methodology, the projects central to the dissertation are not reducible to projects in psychology. The dissertation centrally addresses evaluative issues concerning the nature and conditions of epistemic justification and conditions under which we have knowledge, including moral and evaluative knowledge. These issues go beyond the typical foci of research psychologists. Thus, while the method

used here is broadly naturalistic, the dissertation's epistemological and evaluative aims go beyond 'naturalized epistemology' as W. V. O. Quine understood it.

CHAPTER 1

EMOTION AND THE NEW EPISTEMIC CHALLENGE FROM COGNITIVE PENETRABILITY

0 Introduction

Experiences—visual, emotional, or otherwise—play a role in providing us with justification to believe claims about the world. Some accounts of how experiences provide justification emphasize the role of the experiences' distinctive phenomenology, i.e. 'what it is like' to have the experience. Other accounts emphasize the justificatory role to the experiences' etiology. A number of authors have used cases of cognitively penetrated visual experience (more on these below) to raise an epistemic challenge for theories of perceptual justification that emphasize the justificatory role of phenomenology rather than etiology.[1] Proponents of the challenge argue that cognitively penetrated visual experiences can fail to provide justification because they have improper etiologies. However, extant arguments for the challenge are subject to formidable objections. In this paper, I present the challengers' key claims, raise objections to previous attempts to establish them, and then offer a novel argument in support of the challenge. My argument relies on an analogy between cognitively penetrated visual and emotional experiences. I argue that some emotional experiences fail to provide the relevant justification because of their improper etiologies and conclude that analogous cognitively penetrated visual experiences fail to

provide the relevant justification because of their etiologies, as well.

Although the main aim of the paper is to draw conclusions about justification provided by visual and emotional experiences, my strategy underscores a methodological point as well. The philosophy of perception typically focuses on vision. Recently, coverage has expanded to other perceptual modalities[2] and 'quasi-perceptual' states such as moral and philosophical intuition.[3] The typical strategy is to apply lessons learned from considering vision to a wider range of cases. The approach is largely one-way, taking vision as central and working out toward the periphery. My strategy in the present paper is to go 'in reverse' by arguing for claims in the epistemology of visual experience using premises about emotional states. The argument thereby provides an opposing lane of theoretical traffic in the epistemology of perception, which I hope to show is a welcome addition.

Chapter 1 is organized as follows. In §1, I explain what cognitive penetration is and how it is supposed to raise a challenge for various theories of perceptual justification. In §2, I put the challenge in historical context by distinguishing it from a related debate in the philosophy of science. In §3, I raise objections to three attempts to establish the challenge's key claims. In §4, I show how the notion of cognitive penetration can be defined for emotional experiences, present the argument by analogy with emotion for the challenge's key claims, and respond to potential

[1] Prominent targets of the challenge include Pryor (2000, 2004) and Huemer (2001, 2007).
[2] See, for example, Batty (2010a,b).
[3] See, for example, Huemer (2001), Vayrynen (2006), Chudnoff (2011), and Bengson (2010).

objections. In §5, I show how the argument from emotion avoids the objections to previous attempts to establish the challenge's key claims. I conclude in §6.

1 The New Epistemic Challenge from Cognitive Penetrability

The psychological process of cognitive penetration figures centrally in my discussion. I begin by saying more about it. In other top-down processes, background psychological states such as hopes, fears, and beliefs can cause a subject to shift their spatial attention, change their location, or alter the condition of their own sensory organs (e.g. by pressing on their eyes). In cognitive penetration, however, each of these factors is held fixed.

More precisely, a visual experience is cognitively penetrable with respect to some content or aspect of phenomenal character *c* if and only if two subjects (or the same subject at different times) can differ with respect to whether their experience has *c*, and the difference is the result of a causal process tracing back to a non-visual, psychological state of the subject, where we hold fixed between the two subjects (or one subject at different times) the following: (i) the stimuli impacting their sensory receptors, (ii) the subjects' spatial attention, and (iii) the conditions of the subjects' sensory organs. In the most extreme case, the subject's cognitive states cause the subject to have an entire experience they would not otherwise have had.[4]

[4] For similar definitions, see Siegel (2012), MacPherson (2012), and Stokes (forthcoming). One could define cognitive penetration more narrowly, by adding a further condition stating that the causal penetrating process is semantically

Although 'cognitive penetrability' is defined above for visual experiences, the definition can be extended to other mental states. As I explain below, such states include emotional experiences.

To illustrate cognitive penetration and set up the chapter's main discussion, I present the following example.

> **Angry Looking Jack** Before seeing Jack, Jill has the true but unjustified belief that *Jack is angry*. When she sees him, Jill's unjustified background belief causes her to have a visual experience in which Jack looks angry. If she had lacked the belief, Jack would not have looked angry to her. In addition, Jill is reasonably ignorant of the causal role her background belief plays. As the result of her experience, Jill reaffirms her belief that *Jack is angry* (Siegel, 2012: 2).[5]

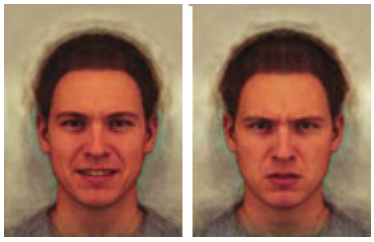
---

significant. See Pylyshyn (1999) for this addition and Stokes (forthcoming) for reasons to resist it. The arguments in the present paper are equally successful whether we include Pylyshyn's semantic condition or not. Alternatively, one could define cognitive penetration more broadly. For example, Lyons (2011) omits condition (ii); as such, cognitive penetration for Lyons can involve shifts in attention in ways in which it cannot on my definition. I use the narrower notion for present purposes, because the key epistemic verdicts I aim to establish in this paper are more plausible for cases involving cognitive penetration in the narrower sense than in Lyons's broader sense. In addition, as I discuss in Chapter 2, I think the shape of the cognitive penetrability challenge is clearer on the narrower definition. In addition, several further points of clarification are in order. Cognitive penetration occurs only when the psychological penetrating state is not part of the penetrated system. Influence by one state within the visual system on another does not count as cognitive penetration; influence by, e.g., a belief state outside the visual system on a visual state can count as cognitive penetration, but only if the other conditions are met. Regarding condition (i): the stimuli impacting the two subjects' sensory receptors must be *qualitatively identical*, but need not be numerically identical. Regarding condition (ii): The relevant differences of spatial attention do not include all differences in attention. That would exclude cases of cognitive penetration where the subject attends to different objects or properties because the process of cognitive penetration causes the experience to represent different objects and properties for the subject to attend to. The idea is to hold fixed where in the phenomenal space the subject attends. Regarding condition (iii): the subject's sensory organs must not be in different conditions due to anything other than the effects of the background states. For example, one does not get a case of cognitive penetration if conditions (i) and (ii) are satisfied, but only one of the subjects has a normally functioning visual cortex due to a brain tumor.

[5] I use Angry Looking Jack in the main text because it figures prominently in the published discussions to which I respond. Paul Ekman (1972, 1993, 2007) has done extensive psychological field work documenting pan-cultural facial expressions of emotion (cf. Ekman & Friesen, 1971, 1986). It is quite plausible that there are stereotypical angry facial

To fix on the phenomenology of Jill's experience, suppose that as a result of cognitive penetration it visually seems to Jill as if Jack's face has the angry expression in Illustration 1b. Had Jill lacked the background belief that Jack is angry, Jack's face would have appeared to Jill with neutral expression in Illustration 1a.

Illustration 1



a b

Illustration 1a depicts a neutral facial expression. Illustration 1b depicts an angry facial expression. Images are from Penton-Voak et al. (forthcoming).

A non-penetrated experience with qualitatively identical phenomenology to that of Illustration 1b would typically provide Jill with sufficient justification to form a justified belief that Jack is angry, absent defeating evidence. This feature of the case is important for how it functions as a putative counterexample to theories of

---

expressions. However, the arguments work at least as well, and perhaps better, if we substitute the following case throughout the discussion.

> **Eye Exam** Earl is at the optometrist taking an eye exam. Before taking the exam, Earl has the unjustified belief that the farthest left letter in the bottom row is a vowel. In fact it is a Q. But Earl's unjustified background belief causes Earl to have an experience as of an O through the process of cognitive penetration.

The Eye Exam case has an even sharper phenomenal contrast between the penetrated and non-penetrated state. It is clearer that visual experiences can have the contents concerning letters than that they can have contents concerning emotional states. And thus it is clearer the dogmatism makes predictions about the case. Thanks for Carl Ginet for suggesting the example.

justification, and I return to it below. We can further suppose that the counterfactual difference in phenomenology and content described in the case would obtain holding fixed the condition of and stimuli impacting Jill's sensory organs and Jill's spatial attention, making Angry Looking Jack a case of cognitive penetration.[6]

A number of authors have raised an epistemic challenge from cases of cognitive penetration like Angry Looking Jack.[7] The intuitive epistemic verdicts about the case are supposed to include the following:

> **Verdict (a)** Jill's cognitively penetrated experience fails to provide her with the degree and kind of *propositional* justification to believe that Jack is angry that she would typically receive from a non-penetrated experience with the same content and phenomenology.

> **Verdict (b)** Jill's post-experience belief that Jack is angry is *doxastically*

[6] As noted above, Angry Looking Jack has featured frequently in recent discussions of cognitive penetration. Such examples may or may not be fictitious. However, there is experimental evidence that cognitive penetration of the sort described in Angry Looking Jack can occur. One of the clearest pieces of evidence comes from a series of experiments by Levin and Banaji (2006). In one of the experiments, participants were shown a gray-scale image of a racially ambiguous face on a computer screen. The face was labeled with either the word 'white' or the word 'black'. The participants then matched the face to a shade of gray. Participants in the group shown images with the 'white' label matched the face to lighter shades of gray than did those in the group with the 'black' label. Levin and Banaji's study provides evidence that cognitive states used to process linguistic data concerning the word 'white' or 'black' on the label can have an influence on the phenomenology of the participants' visual experiences. Since the states used to process the linguistic data are presumably outside the visual system, the results cannot easily be explained as arising due to intra-systematic processes. And since the differences in the phenomenal character of the participants' experiences pertained to shades of gray, it is implausible that the differences were due to shifts in stimuli, attention, or the condition of the participants' sensory organs. For a thorough discussion of the evidence of cognitive penetration from this and related experiments, see MacPherson (2012). For discussion of evidence from experiments involving the role of background desires, see Stokes (2012). For responses to some of the central arguments for impenetrability in Fodor (1983), see Prinz (2006).

unjustified.[8]

Discussions of the challenge from cognitive penetration have so far focused on how the challenge arises for theories of perceptual justification that deny that the etiology of an experience has a justification-related role to play in addition to fixing the experience's content and phenomenology and in grounding defeaters—where a defeater for some justification *j* is understood to be, roughly, evidence one possesses that undermines *j*. That is, proponents of the challenge target accounts of perceptual justification inconsistent with the following claim:

> **Etiological thesis** A visual experience as of *p* can fail to provide S with the usual degree and kind of justification for *p* as a result of the experience's having an improper etiology, without the etiology also providing a defeater for the relevant justification.[9]

Angry Looking Jack is supposed to be a counterexample to views that are inconsistent with the etiological thesis, such as dogmatism.

---

[7] Proponents of the challenge include Markie (2005, 2006), Goldman (2008a), Siegel (2012, forthcoming), Lyons (2011), Jackson (2011), and McGrath (forthcoming-a, forthcoming-b).

[8] Because the cognitive penetrability challenge raises questions about the nature of epistemic justification and the conditions under which it is provided, it is best not to prejudge the debate by offering precise accounts of propositional and doxastic justification. However, the rough ideas are as follows. S has propositional justification for *p* if and only if (and to the degree that) S has good epistemic reason to believe *p*. And S's belief that *p* is doxastically justified if and only if S believes *p* for good epistemic reason.

Dogmatism A visual experience as of *p* provides some defeasible, immediate propositional justification for *p* in virtue of the experience's having a distinctive phenomenology with respect to *p*.[10]

Angry Looking Jack is supposed to be a counterexample to dogmatism as follows. Jill has a visual experience as of Jack's being angry, so dogmatism predicts that Jill receives some defeasible, immediate, propositional justification for the claim that Jack is angry. Moreover, dogmatism predicts that Jill's penetrated experience provides her with the degree of defeasible, immediate, propositional justification for the claim that Jack is angry that would usually be provided by a non-penetrated experience with the same phenomenology and content. So dogmatism predicts that Jill receives the usual propositional justification for the claim that Jack is angry, contrary to verdict (a). In addition, such justification is typically sufficient for an outright belief that Jack is angry. And, since Jill is reasonably unaware of her penetrating belief's causal role, she lacks a defeater for any justification she might receive from the experience

2 The New Challenge and an Older Debate about Theory-Ladenness

The new epistemic challenge from cognitive penetrability has important connections

[9] Note that the formulation in the text implies that visual experiences can share contents with beliefs. On such an assumption, Pylyshyn's semantic condition can be satisfied. However, the argument can be formulated with both the assumption and Pylyshyn's condition omitted.

[10] Huemer's phenomenal conservatism extends dogmatist principles to non-visual 'seemings' states.

to an older debate in the philosophy of science about the theory-ladenness of observation. The evidence for the cognitive penetrability of perception originates with the New Look psychologists, whose research program led to numerous studies of top-down effects on perception from the 1940s onward.[11] In the 1980s, Jerry Fodor mounted a significant attack on the New Look program and against the tradition of the so-called "Harvard relativists"[12] who used considerations from the sociology and philosophy of science to argue against the objectivity of perceptual experience and observation. Fodor (1983, 1984, 1988) argued for a much smaller role for top-down processing in perception; instead, he emphasized the role of bottom-up, informationally encapsulated mental modules.[13]

One might argue for the importance of the new challenge from cognitive penetration by claiming that it breathes new life into the older debate about theory-ladenness in the philosophy of science. Alternatively, one might argue for the unimportance of the new challenge by claiming that it merely rehashes the older debate. However, neither response to the challenge is correct. It is more accurate to describe the recent challenge as raising a new debate in its own right. There are at least two important differences between the debates.

First, the cognitive scientific aspect of Fodor's debate was about the frequency

[11] For further discussion of the New Look psychologists results, see Chapter 2.
[12] See Goodman (1978), Kuhn (1962) and others cited in Fodor (1984).
[13] See Pylyshyn (1980, 1984, 1999) and Raftopoulos (2001) for further arguments against New Look psychology. Like Fodor, Pylyshyn concedes that perception is cognitively penetrable in general; he defends only the claim that early visual processes are impenetrable.

with which cognitive penetration occurs, whereas the recent debate is not. Fodor accepted that perception is cognitively penetrable to at least a minimal degree.[14] Against the New Look psychologists, Fodor maintained that such top-down influence was the exception rather than the rule. By contrast, the new challenge from cognitive penetration gets off the ground with only a few select cases of cognitive penetration. Since these cases are supposed to serve only as counterexamples to theories of perceptual justification, it is not important for the debate whether the examples occur frequently. In fact, to the extent that the views in question entail claims about merely possible cases of cognitive penetration, it is not even essential that the cases actually occur; though as noted above, it is plausible that they do.[15]

Second, the epistemic aspect of Fodor's debate with the New Look psychologists and Harvard relativists concerned the objectivity of perception and its suitability as a foundation for scientific consensus and knowledge. For Fodor,

> To get from a cognitivist interpretation of perception to *any epistemologically interesting* version of the conclusion that observation is theory dependent, you need not only the premise that perception is problem solving, but also the premise that perceptual problem solving has access to ALL (or, anyhow,

[14] For example, Fodor writes, "The point of perception is the fixation of belief, and the fixation of belief is a *conservative* process—one that is sensitive, in a variety of ways, to what the perceiver already knows. Input analyses may be informationally encapsulated, but perception surely is not." (1983: 73, emphasis original)

[15] The last point is related to a further difference: experimental evidence for more recent claims concerning cognitive penetration is arguably better than that provided by the New Look psychologists. See Chapter 2 for further discussion.

arbitrarily much) of the background information at the perceiver's disposal. (1984: 35, italics added; capitalization original.)

Fodor focuses on providing evidence against the second premise above, the one that says that perceptual systems have access to arbitrarily much of the subject's background information. The new challenge from cognitive penetration is compatible with Fodor's denial of that extreme claim and is not focused primarily on the objectivity of science. Rather, the new challenge is concerned with specific verdicts concerning a subject's justification in cases where these background top-down processes have an effect on perception, however often that is. By drawing attention to possible counterexamples to theories of perceptual justification, the new challenge also suggests epistemologically interesting claims related to theory-dependence that Fodor overlooks.

In sum, although the new challenge has roots in a historically important debate about the theory-ladenness of observation, the new challenge is orthogonal to the more contentious issues that were central to the older debate. To accept the key claims in the new challenge one need not be a relativist about confirmation in science and one need not think that cognition relentlessly shapes perceptual processing.. So, since the defeasible, propositional justification that dogmatism predicts Jill receives from her experience is sufficient for an outright belief that Jack is angry, Jill bases her reaffirmed belief on the experience in the usual way, and the justification is not

defeated, dogmatists are under pressure to accept that Jill's post-experience belief is doxastically justified, contrary to verdict (b).

3 Problems for Recent Attempts to Establish the Key Verdicts

The challenge from cognitive penetration poses a threat to theories of justification only if the key verdicts are established. In this section, I raise objections to three attempts to establish them. I do so for a specific purpose: in subsequent sections I show that the argument by analogy with emotion avoids the objections.

3.1 Appeals to Intuition

One might attempt to establish the key verdicts by appeal to intuition. Perhaps it is intuitively obvious that Jill's experience fails to give her the usual propositional and doxastic justification to believe that Jack is angry. However, even the verdicts are intuitive, they are not obvious, and appeals to intuition are insufficient to establish them. Everyone in the debate can agree that *something* is defective about Jill's post-experience belief that Jack is angry, but it is not clear that the defect pertains to justification. The intuitive defect could be Jill's lack of knowledge.[16] As Gettier (1963)

[16] See Tucker (2010) for a response along these lines. Siegel (forthcoming) and Markie (p.c. cited in Tucker, 2010) both give arguments aimed at ruling out the claim that the intuitions concern lack of knowledge. I think the arguments are unsuccessful. The strategy in each case is to try to isolate the justification intuition by arguing that there is some defect in addition to the subject's lack of knowledge, and then to claim that that defect pertains to justification. Although I won't enter into details, a problem for both strategies is that there may be multiple defects which pertain to lack of knowledge without pertaining to justification. Both authors seem to assume that once they have identified a defect pertaining to knowledge, any other defects must pertain to other epistemic properties, such as justification. But this is

argued, justified true beliefs can fail to be knowledge. Depending on how one fills in the details of the case, Jill might fail to know that Jack is angry, even though he is angry. Additionally, the defect might pertain to Jill's epistemic irresponsibility or epistemic vice, where these are not equivalent to defects of epistemic justification.[17]

There is an additional argument that establishing the key verdicts requires more than appeal to intuition. Markie (2005) and Siegel (2012) report that they have the intuition that subjects like Jill fail to get any justification from their experience, including any defeasible, propositional justification. However, it is implausible that Jill fails to get any defeasible, propositional justification from her penetrated experience. Consider the following claims.

(1) It visually appears to me as if Jack is angry.

(2) If it visually appears to me as if Jack is angry, then probably Jack is angry.

(3) So, probably Jack is angry.

Jill has propositional justification for (1) as the result of having her experience as of Jack's being angry. Even if there is something defective about the experience or its etiology, Jill gets justification to believe that it appears to her as if Jack is angry. Jill also possesses propositional justification for (2). Typical subjects possess extensive evidence of the reliability of their own visual experiences. If we suppose that Jill is a

not the case. A belief might fail to be knowledge for more than one reason, none of which pertains to a failure of justification.

[17] See McGrath (forthcoming-a) for this suggestion.

typical subject, she has strong positive evidence that her belief was reliably produced. Moreover, given that the case must be set up such that she lacks defeaters for any justification she receives from her experience for the claim that Jack is angry, she lacks undercutting defeating evidence against reliability in the particular case at issue. So, Jill has justification for (2). Finally, there is nothing suspicious about an inference or epistemic support relation from the conjunction of (1) and (2) to (3). The claims form a valid syllogism and an inference from (1) and (2) to (3) is reasonable. So, Jill gets some propositional justification for the claim that Jack is angry via her justification for (1) and (2).[18] The conclusion contradicts Markie's and Siegel's reported intuitions about the case, so we have further reason to believe that their intuitions are misguided or that they are misreporting the content of their intuitions. In either case, more than an appeal to intuition is needed to establish the key verdicts.

3.2 The Testimony Analogy

Susanna Siegel (2012) suggests that cognitive penetration is like a case of testimony. She presents the testimony analogy as a problem for dogmatism. Comparing Angry Looking Jack to a tight gossip circle Siegel writes,

> We can compare this situation to a gossip circle. In a gossip circle, Jill tells Jack

[18] The above considerations about a possible introspective route to justification in the case suggest that the relevant justification to focus on is non-introspective. This suggests that the challenge from cognitive penetration does not arise for Crispin Wright's (2007) view, *pace* Siegel (2012). Moreover, I should note that, in order to establish verdicts (a) and (b), I must assume that subjects in the relevant cases form their beliefs

> that *p*, Jack believes her but quickly forgets that she's the source of his belief. Shortly afterward, Jack tells Jill that *p*. It seems silly for Jill to take Jack's report that *p* as providing much if any additional support for *p*, beyond whatever evidence she already had. On the face of it, this looks like a feedback loop in which no new justification is introduced. Similarly, when beliefs are formed on the basis of cognitively penetrated experience, it is as if your belief that *p* told you to have an experience that *p*, and then your experience that *p* told you to believe that *p*. (Siegel, 2012: 2)

Siegel suggests that Jill fails to get additional justification for her belief that *p* as the result of Jack's testimony. Siegel then contends that, by analogy, Jill fails to get additional justification from her penetrated experience for her belief that Jack is angry. In both cases, Siegel claims, there is a "pernicious" and "ridiculous" feedback loop which prevents Jill from getting the usual justification from her experience.

However, it seems to me that the testimony analogy does not support the key verdicts, and may undermine them. In the gossip circle case, Siegel tells us that *Jack* quickly forgets that Jill is his source for *p* before he quickly reports back to Jill that *p*. One naturally imagines that when Jack reports back, Jill remembers that she is Jack's source of information about *p* even though Jack has forgotten. In this version of the example, Jill cannot rationally raise her credence in *p*, and it may seem plausible that Jill's experience fails to give her additional justification for *p*. But this version is not

relevantly analogous to Angry Looking Jack. Gossip Jill clearly has a defeater for any justification she might have received from Jack's testimony, since she remembers that he is merely repeating back to her what she just told him. By contrast, in Angry Looking Jack, Jill lacks such a defeater since she is reasonably unaware of the penetration.

A version of the gossip circle example analogous to Angry Looking Jack entails that Jill is reasonably unaware that the person she just told that *p* is now repeating it back to her. For Angry Looking Jack to serve as a counterexample to dogmatism, Jill must lack defeaters for any justification she receives from her penetrated experience. If no such entailment holds, the presence of defeaters, rather than the improper etiology, could explain Jill's lack of justification and dogmatism would not make any false predictions about the case.

Here is a reformulated version of the case with the proper structure.

> **Disguised Gossip Circle** Jill initially believes without justification that (*p*) there's a party at Pat's. Jill tells Jack that there's a party at Pat's. Jack then convincingly disguises himself as Fred, a friend of Jill's whom she does not realize Jack has ever met. Jill then sees Jack, disguised as Fred, and justifiedly thinks he's Fred, whom she reasonably takes to be a reliable source of information independent of Jack. Jack-in-disguise then tells Jill that there is a party at Pat's. Jill reaffirms that *p* on the basis of the testimony.

In Disguised Gossip Circle, Jill is reasonably unaware that she is receiving testimony from a source that originates with her, and she has excellent reason to think that the source is credible. However, regarding the reformulated example, I do not have the intuition that Jill has a doxastically unjustified belief that *p* based on the relevant testimony. In fact, I think Jill's testimonial belief is justified in the reformulated case. If the Disguised Gossip Circle is a good analogy for the sort of process present in a case of cognitive penetration like Angry Looking Jack, the example does not support—and may undermine—the key verdicts.

The analogy between a gossip circle and cases of cognitive penetration of visual experience has historical support. There is a long tradition of using "the testimony of the senses" as a metaphor of sensory perception.[19] It is, I think, usually assumed that the senses (*qua* testifiers) are independent of the subject's antecedent beliefs and can provide largely unbiased testimony about the world. However, as the gossip circle cases suggest, the analogy could be extended to cover feedback loops like those found in cognitive penetration, since testimony chains can include gossip circles originating with the subject.

However, although the metaphor is venerable, it is not clear that Disguised Gossip Circle is analogous to Angry Looking Jack in the relevant respects. Most

[19] For example, Descartes, Hume, and Reid all employ the metaphor. Contemporary epistemologists do as well, including externalists such as Plantinga and Goldman.

epistemologists agree that there is an important difference between justificatory processes that are contained within one's mental system (e.g. inferences) and those that are not (e.g. testimony chains). Cognitive penetration proceeds within a subject's mental system. By contrast, in the Disguised Gossip Circle, the causal process from Jill's background belief to her reaffirmation of *p* traces outside her mental system. This provides the potential for a relevant difference between Disguised Gossip Circle and Angry Looking Jack.

In sum, the testimony analogy fails to support the cognitive penetrability challenge in at least one of the following two ways. First, the testimony examples that are most similar to Angry Looking Jack do not support the key verdicts and may support their negations. Second, the analogy may be flawed in any case, due to differences in the relevant causal processes. The upshot is that the cognitive penetrability challenge cannot be supported by the testimony analogy.

3.3 The Belief Analogy

Siegel (forthcoming) aims to support the key verdicts in the cognitive penetrability challenge with an argument by analogy with belief. She compares the etiology of Jill's cognitively penetrated visual experience with the etiology of an unjustified belief. Here is a reconstruction of the argument as applied to Angry Looking Jack.[20]

[20] In more recent work, Siegel also compares the etiologies of experiences to the etiologies of beliefs caused by wishful thinking. The analogy in those cases may be more promising than the analogy to cases of belief preservation. However, my focus is on the penetration by a belief in Angry Looking Jack, so I assess the belief analogy argument as applied to it.

The Argument by Analogy with Belief

B1. The etiology, X, of Jill's cognitively penetrated experience as of Jack's angry face is similar to the etiology, X*, of an unjustified belief.

B2. A belief with etiology X* is unjustified in virtue of its causal history, and thereby fails to provide any relevant justification.

C. So, the key verdicts are true for the cases of cognitive penetration of visual experience, including Angry Looking Jack.

The argument is not deductively valid. It relies on the strength of the analogy between the etiologies of a belief and of a cognitively penetrated visual experience.

Premise B2 is plausible. It entails that a belief with etiology X* is unjustified in virtue of its causal history, and thereby fails to provide the usual justification. Siegel proposes that the relevant etiology for comparison is of belief preservation of an originally unjustified belief.[21] Suppose that you form an unjustified belief. As time passes, you retain the belief but forget how you formed it. Siegel suggests that the intuitively correct verdict for such a case is that your belief at the later time is doxastically unjustified, and that the improper etiology of the belief is what makes it unjustified. In general, the etiology of a belief is relevant to whether the belief is

[21] An alternative suggestion is to compare the cognitive penetration in Angry Looking Jack to an inference from an unjustified belief that *p* to the conclusion that *p*. However, it is difficult to see how one could make such an inference

doxastically justified. For example, a belief that *q* could be unjustified in virtue of being grounded on an unjustified belief that *p*.

B1 is less plausible than B2. We can grant that in the belief preservation case, your later belief is unjustified due to its etiology. We should still raise doubts about the claim that belief preservation is analogous to cognitive penetration. The brain processes involved in each case are very different. Belief preservation is a largely static process. Cognitive penetration requires neural activity in various parts of the brain, including the visual cortex where the experience is at least partly realized. There is no straightforward inference from differences in neurology to epistemic differences; however, it is not clear why we should treat such different processes as relevantly similar for Siegel's epistemic purposes. In general, beliefs and visual experiences are very different sorts of state. As Siegel (forthcoming) admits, "You wouldn't expect a sausage machine to make M&M candy using the very same mechanism." Unfortunately, Siegel provides little argument that the two etiologies are relevantly similar.

In addition to the worry that the etiologies of Jill's experience and a preserved belief are not sufficiently similar, there is a more serious worry about the belief analogy argument. The worry targets the inference from the conjunction of B1 and B2 to the conclusion C. Even if we grant that the etiology of a belief and a penetrated

unconsciously, as would be required by the no-defeaters set-up of the case. The notion of inferring a claim from itself is strained as it is. The best way to make sense of it is as a conscious process of deliberately making a logical point.

experience can have similar etiologies, we should be hesitant to draw Siegel's conclusion about their respective statuses as justifiers. Visual experiences have a distinctive kind of phenomenology which plausibly plays a role in providing the relevant justification (more on the relevant phenomenology in the next section). It is controversial whether belief states have any phenomenology at all. However, I know of no one who argues that beliefs have the kind of phenomenology experiences have.[22] These differences in phenomenology provide a relevant difference between visual states and beliefs. One could argue that, unlike beliefs, visual experiences can provide the usual justification regardless of their etiologies, because they have a kind of phenomenology that beliefs lack. Indeed, as noted earlier, dogmatism and views like it suggest that phenomenology can play just such a role. The belief analogy provides no answer to this response.

4 An Argument by Analogy with Cases of Emotion

To move beyond relying on controversial intuitions and to avoid the problems for the analogies to testimony and belief, proponents of the cognitive penetrability challenge need a new argument to establish the key verdicts. In this section, I develop an argument from analogy with cases of emotion. The emotion analogy argument has the following structure.

[22] One possible exception here might be putatively self-evident beliefs or claims. However, the beliefs at issue in the present discussion are not candidates for self-evidence beliefs or claims.

The Argument by Analogy with Emotion

P1. There are cases of cognitively penetrated emotional states that are analogous to Angry Looking Jack and other cases of cognitively penetrated visual experience.

P2. The relevant verdicts are true for the cases of cognitively penetrated emotion.

C. So, the key verdicts are true for the cases of cognitive penetration of visual experience, including Angry Looking Jack.

The argument is not deductively valid. C is supposed to be inferable from P1 and P2 on the strength of the analogy between the cases of cognitive penetration of visual and emotional experience. In the next subsection, I argue that some cases of emotion are relevantly analogous to some visual cases, like Angry Looking Jack (P1). In the following subsection, I argue for the verdicts about justification for the relevant cases of emotion (P2). I conclude the section by addressing two objections.

4.1 Establishing the Analogy (and P1)

In this sub-section, I argue that there are examples of emotional states that are relevantly analogous to some visual states in their phenomenology, justificatory role, and etiology. I begin with the following example.

Fear of Snakes Sara sees a snake along a path, hissing and rattling at her. In response, Sara has an affective experience of fear, by which she consciously senses that there is danger. She comes to believe that she is in danger on the basis of her feeling of fear.

Sara's feeling of fear is analogous to a perceptual state in several ways. The first point of analogy concerns the phenomenology of such states. Like a perceptual state, a fear state has a distinctive phenomenology. That it does can be supported by introspection; one can be introspectively aware of the phenomenology of one's fear state. That fear has a distinctive phenomenology can also be supported by using neurological and biochemical evidence indicating that there are distinct autonomic nervous system (ANS) arousal patterns for various emotions, which correlate with distinct kinds of phenomenology.[23]

Fear states share a general phenomenal characteristic with visual states: both have what I will call 'presentational phenomenology'. When one feels fear it often feels to one as if a danger is being presented to one. Such states have a phenomenal force as of something's being revealed to one. Likewise, when it visually appears to one as if *p*, it feels to one as if the fact that *p* is revealed to one. In this respect, the

[23] For a summary of the relevant data, see Oatley, Keltner, & Jenkins (2006), Ch. 5. See also the data reported by Griffiths (1997: 81-3) and Prinz (2004: 69-71). Griffiths and Prinz also provide compelling evidence in response to

phenomenology of Sara's fear is analogous to the phenomenology of Jill's visual experience.[24]

Another point of analogy concerns the justificatory roles of such states. Sara's fear plays an epistemic justifying role similar to the justifying role played by perceptual states. Sara does not come to merely believe that she is in danger; she comes to justifiedly believe it on the basis of her emotional experience. Emotions can serve as good grounds for judgments about one's environment. Thus, as with visual states, at least some emotional states can play a role in justifying beliefs some of the time.[25]

We now have points of analogy between some emotional states and visual states: both can have distinctive, presentational phenomenology, and both can sometimes provide justification for beliefs. However, we do not yet have cases of emotion that are properly analogous to Angry Looking Jack. To get such cases, we extend the concept of cognitive penetrability to emotional states as follows.

---

experiments like in Schacter & Singer (1962), which some have interpreted as undermining the claim of distinct phenomenology.

[24] One might resist the claim that emotional states have presentational phenomenology. To do so, one might offer a precise characterization of the phenomenology and argue that emotions lack the phenomenology so characterized. However, it is difficult to say precisely what presentational phenomenology is, and there is disagreement in the literature concerning how best to characterize it. A contrast is often drawn between perceptual and memory states (which have presentational phenomenology) and imagination and belief states (which do not). To raise the cognitive penetrability challenge, it is enough that the cases of emotion under discussion have a sort of phenomenology that can plausibly be called presentational, and they do have such a phenomenology. For further discussion of the relevant phenomenology under various descriptions, see Heck (2000), Pryor (2000, 2004), Huemer (2001, 2007), and Chudnoff (2011). For additional arguments that affective states have the relevant phenomenology, see Johnston (2001).

[25] The argument in the main text implies that emotions can sometimes provide the relevant justification, and this is the position I endorse. However, that assumption could perhaps be dropped. If emotions never provide justification for beliefs, it might well be for reasons that support the etiological thesis and the key verdicts, rather than undermining them. For example, some readers may think that fear and disgust are (or require) judgments. Such a view need not undermine the argument from emotion. It would rather provide a different route to the same conclusion. If Sara's affective fear state as of danger cannot justify the belief in danger that causes it, the challenge arises: for the view provides reason to think that the belief's etiological role is what explains the state's failure to provide the relevant

Emotional states are cognitively penetrable with respect to some content or aspect of phenomenal character *c* if and only if two subjects (or the same subject at different times) can differ with respect to whether their emotional states have *c* as the result of a causal process tracing back to a cognitive state of the subject, holding fixed between the two subjects (i) the stimuli impacting their sensory receptors, (ii) the subjects' spatial attention, and (iii) the conditions of the subjects' sensory and emotional organs. As with visual experience, in some cases of cognitively penetrated emotion, a subject's cognitive states cause them to have an emotional experience they would not otherwise have had. To illustrate cognitively penetrated emotion, I present the following case.

> **Whistle Fear** You believe that you are alone in the house late at night, when you hear someone whistle close behind you. Your belief that you are alone (together with the auditory stimuli) plays a role in causing you to feel fear. If you lacked the belief, you would not feel fear in response to those stimuli.[26]

Suppose that, had you not believed that you were alone, you would have believed that your spouse was with you in the house. In otherwise identical circumstances, upon receiving the same external stimuli you would have directed your attention to the

---

justification. By analogy, one could then argue, the role of Jill's belief also prevents her visual state from playing the relevant justifying role.

same location, you would have been located in the same place, and your sensory and emotional organs would have been in the same condition. But, believing that your spouse was home, you would not have felt fear. With these details in place, Whistle Fear is a good candidate for a cognitively penetrated emotional experience.

One might worry that in Whistle Fear the subject does not respond to the same distal stimuli that they would have if they had lacked the influential background belief. So, one might worry that it is not a case of cognitive penetration as defined.[27] For example, the subject's differing bodily states could be counted as different distal stimuli. A tradition in emotion theory dating to William James (1884) and Carl Lange (1885) supports the idea. According to the James-Lange theory, emotions are perceptions of one's own bodily states. If one feels fear, one's bodily states would likely be different than if one did not feel fear, in which case the distal (bodily) stimuli are not held fixed. So, if we count the different bodily sensations as among the distal stimuli the subject perceives, Whistle Fear is not a case of cognitive penetration.[28]

There are at least two responses to the worry. One option is to argue that bodily states are not among the distal stimuli. Perhaps subjects use bodily states to process the distal stimuli that inform their emotional experiences, but the bodily states are not included among the distal stimuli. The response requires denying the James-

---

[26] The example is drawn from Pizarro and Bloom (2001).

[27] It is easier to establish that there are cases of cognitively penetrated emotion on a broader definition of 'cognitive penetrability' such as that is Lyons (2011), discussed in note 4 above.

Lange claim that emotions are perceptions *of* bodily states. However, it is compatible with a modified account on which the awareness of bodily sensations is an important part of emotional processing.

The second response avoids settling whether bodily states are distal stimuli. It focuses on cases where an emotion arises without a change in the relevant bodily states. There is empirical evidence that changes in emotion phenomenology can occur without bodily changes. For example, Hohmann (1966) found that patients with greatly diminished capacity to sense their own bodily states due to spinal cord injuries still feel robust emotion phenomenology similar to what they experienced prior to their injuries. The results suggest that, even if awareness of bodily changes is often involved in emotional experience, the existence of the changes is not necessary for one to feel the relevant emotion. In addition, Stemmler (1989) found that participants who self-induced emotional states by recalling an episode of fear or anger experienced distinct emotions without distinct physiological reactions—as measured by, e.g., heart rate, head temperature, and skin conductance.[29] The results suggest that patients can have distinct emotional phenomenology as the result of differences in their cognitive background states and other parts of the brain responsible for emotion processing (e.g. parts of the limbic system), without distinct arousal patterns and while other relevant factors are held fixed. In short, even if emotional phenomenology often

[28] The James-Lange theory is controversial. In particular, it is controversial whether emotions consist *only* in perceptions of bodily states. It is worth clarifying that the above objection requires only that emotions at least partly involve perceptions (or sensations) of bodily states as distal stimuli.

correlates with distinct bodily states, it does not always do so.[30]

Having extended the concept of cognitive penetration to emotional states and having argued that cognitive penetration of emotion can occur, I now provide support for P1 by presenting an example of cognitively penetrated emotional states analogous to Angry Looking Jack. Consider the following case.

> **Fear of Foreigners** Xena holds an unjustified belief that *foreigners are dangerous*. Xena does not interact with foreigners often, and her belief is inactive for a long time. One day she sees a foreigner. Xena's unjustified background belief that *foreigners are dangerous* causes her to experience fear of the person. She is reasonably unaware that her background belief has the relevant causal influence and that the fear is partly derived from the person's looking foreign. As the result of her emotional experience, Xena affirms that the person is dangerous.

The details of the example are realistic. Subjects can be reasonably unaware of the penetrating states that partly cause their emotions. Likewise, subjects can be reasonably unaware of the features of a situation the result of which they feel an emotion.

---

[29] For a review of the relevant literature, see Oatley, Keltner, and Jenkins (2006), Ch. 5.

[30] A further available response requires defining 'cognitive penetrability' more broadly, along the lines of Lyons (2011); see note 4. On such a definition, cognitive penetration is compatible with shifts in spatial attention and distal stimuli, so long as these shifts are caused by cognitive states, perhaps with some further restrictions. However, as explained above, I think that cases of cognitive penetration in the sense defined in the main text provide clearer cases of the relevant

To bring out the relevant points of analogy between Fear of Foreigners and Angry Looking Jack, we can describe the features they have in common. In both cases, an unjustified background belief cognitively penetrates some state of the subject and causes the subject to have the relevant experience. The states both have a distinctive, presentational phenomenology. They both are the sorts of state that can, under some circumstances, provide justification for the subject's beliefs. In addition, the etiologies of the cognitive penetration processes are plausibly analogous: they both proceed from a background belief to a non-belief state with presentational phenomenology, and the subjects are unaware of the causal roles the beliefs play in both cases. Finally, since Xena is reasonably unaware of the role of the background belief, she lacks defeaters for any justification that might be provided by her feeling of fear, just as Jill plausibly lacks defeaters for her post-experience belief that Jack is angry.[31]

The above considerations support P1 in the emotion analogy argument. They indicate that there are cases of cognitively penetrated emotional states that are analogous in relevant respects to Angry Looking Jack.

4.2 Establishing the Key Verdicts for Cases of Emotion (and P2)

In this section, I support P2 by arguing for the relevant verdicts in Fear of Foreigners.

---

phenomenon and the shape of the challenge from cognitive penetrability, so I prefer to work with the narrower definition.

[31] As before, similar examples can be constructed using disgust.

That is, I argue that Xena's cognitively penetrated emotional state fails to provide her with the degree and kind of propositional justification that she would typically receive from a non-penetrated experience with the same content and phenomenology and that Xena's post-experience belief formed on the basis of the experience is doxastically unjustified.

Emotions can be irrational or unjustified.[32] The senses of irrationality or unjustifiedness (I use these terms interchangeably) that are predicable of emotions include a sense of epistemic irrationality that is analogous to the epistemic irrationality of belief. For example, a fear can be irrational or unjustified when the subject knows that the object of his fear is not dangerous. Anger can be unjustified when one possesses strong evidence against an offense's occurrence. For present purposes, it is important that emotions can also be unjustified in virtue of being caused in certain ways, for example, when they are caused by unjustified background beliefs. Fear of Foreigners presents an unjustified emotion of this sort.[33] Xena's unjustified background belief that foreigners are dangerous causes her to feel fear of the person she encounters. She is unaware that her background belief plays a causal role and does not realize that she is afraid of the person because they look foreign to her. Nevertheless, because her fear is caused by an epistemically unjustified background belief in the relevant way, her fear is epistemically unjustified.

[32] For some of the many examples, see Brady (2007, 2009), Deonna and Teroni (2012), Helm (2007), Greenspan (1988), and Pitcher (1965).

I now use the claim that Xena's emotional state is epistemically unjustified to argue for the key verdicts as applied to Fear of Foreigners. The standard view in epistemology about the conditions under which beliefs provide justification is that an epistemically unjustified belief fails to provide any defeasible justification for the relevant claims.[34] The view is widely accepted. It is required for all of the standard solutions to the 'problem of the regress of justification'. For any justified belief *b* one can ask 'In virtue of what is *b* justified?' If *b* is based on some other belief *b** and justified in virtue of its being so based, *b** must also be justified: an unjustified belief cannot serve as the basis for a justified belief. The regress arises because one then has to ask in virtue of what *b** is justified. It is possible to avoid the regress by arguing that an unjustified belief can provide the relevant justification, but none of the standard solutions to the regress problem take this route. The thought common to them all is that an unjustified belief cannot justify other relevant beliefs. There is nothing about the thought that is peculiar to beliefs. The same points apply to judgments and other states. On a plausible extension, it applies to emotions as well. An emotion's status as unjustified prevents it from providing the relevant justification just as a belief's status as unjustified prevents it providing the relevant justification.

I am now in a position to establish the key verdicts for Fear of Foreigners and thereby establish P2 in the argument from emotion. Xena's fear is epistemically

[33] For similar cases used as examples of irrational or unjustified emotions, see Pitcher (1965) and Taylor (1975); cf. Deonna and Teroni (2012: Ch. 8).

unjustified in virtue of its etiology. As such, it is analogous to an epistemically unjustified belief. Moreover, and also analogously with belief, Xena's fear fails to provide her with the degree and kind of propositional justification she would typically receive from a justified fear state with that content and phenomenology. And Xena's post-experience belief that the person is dangerous is doxastically unjustified.

The argument above supports P2 in the emotion analogy argument: the relevant verdicts are true for the cases of cognitively penetrated emotion. Taking P2 together with P1—the claim that the cases are relevantly analogous to Angry Looking Jack—we can conclude that the relevant verdicts are true for Angry Looking Jack.

4.3 Objections and Replies

In this section, I address two potential objections to the emotion analogy argument. First, one might think that the argument relies on a perceptual theory of emotion or something close to that, according to which emotional states are perceptual states. Perceptual theories of emotion are controversial. If the argument relies on a perceptual theory of emotion, accepting the argument comes with significant costs.

In response to the first objection, the argument does not require the claim that emotional states are perceptual states. Nor does it require that emotional states not discussed here are analogous to perceptual states. Perceptual theories of emotion have

[34] By using 'the relevant claims' I intend to exclude introspective justification for claims about one's own beliefs. See section 3.1 for related discussion.

enjoyed a modest recent rise in popularity.[35] The plausibility of such views could be used to provide further support for the argument. However, the argument in no way depends on the success of such theories. Indeed, the argument from emotion presented here does not require that there is a true, unified 'theory of emotions' at all. A common complaint about theories of emotion (including perceptual theories) is that they emphasize a core set of favored examples and then generalize from that core to claims about all emotions, while ignoring a wide range of differences between the core examples and the full range of emotions.[36] My argument in this section faces no such worry. The argument focuses on the core set of examples that tend to motivate perceptual theories, but it does not require generalizing from them to conclusions about other emotions. Instead, the argument involves inferences from core cases of emotion to analogous cases of perceptual experience. Moreover, the argument utilizes claims about emotions that serve as *desiderata* for theorizing about emotion, not by employing a controversial theory of emotion.

Here is a second objection to the emotion analogy argument. An important piece of evidence that the emotional state used in the argument does not provide the relevant justification is that emotional states can have epistemically improper etiologies. And an important piece of evidence for that claim is that emotional states can be assessed as epistemically unjustified in virtue of their improper etiologies. The

[35] For philosophical defenses of perceptual theories of emotion, see for example Clarke (1986), de Sousa (1987), Prinz (2004), Roberts (2003), and Tappolet (2006). Cf. Damasio (1994) and LeDoux (1996), for psychological views that fit well with what could be called a broadly perceptual theory of emotion.

rational assessment of emotional states may seem to provide a potential relevant difference between emotional states and visual states. Visual states do not seem to be rationally assessable. If the argument by analogy with emotion is to succeed we need an explanation for how we can draw the conclusion about visual states despite the apparent difference in rational assessability between the two types of state.

In response to the second objection, it is important to note that the proponent of the argument from emotion need not rely on the claim that visual states are rationally assessable. One could define a notion of (proto)-rationality for visual states.[37] But that is not the version of the argument that I have developed here. Such a move would have to be motivated by general considerations about the nature of visual states or of an important subset of such states. The emotion argument does not require taking a stand on these general issues. Instead, the argument exploits similarities between some cases of emotion and vision without any attempt to generalize the analogy for other pairs of emotional and visual states. The examples of cognitively penetrated visual states discussed here may not be sufficiently representative of the class of visual states as a whole to motivate including such states in the set of rationally assessable states.

The essential point for present purposes is that the cognitively penetrated fear and visual states are sufficiently alike to support the key verdicts. In developing the

---

[36] See Griffiths (1997) for the most thorough development of this line of thought.
[37] Siegel (forthcoming) also notes the availability of this option.

analogy, I have given reasons to think that the target emotional and visual states have analogous etiologies, phenomenology, and justificatory roles. These analogous characteristics obtain for Jill's and Xena's visual and emotional states, respectively. Both states have etiologies involving cognitive penetration by an unjustified background belief. Both states have a similar sort of presentational phenomenology. And visual and fear states can play similar justificatory roles when cognitive penetration is not involved. These points of analogy support drawing the same epistemic verdicts in both cases. Since Xena's state fails to provide the relevant justification, we can conclude that Jill's state fails to provide the relevant justification, as well.

5 Advantages of the Emotion Analogy Argument

In the remainder of the paper, I show how the argument from emotion solves the problems for other attempts to establish the key verdicts in the challenge from cognitive penetrability.

5.1 Appeals to Intuitions Revisited

In §3.1, I discussed attempts to establish the key verdicts by appeal to intuitions about cases of cognitively penetrated visual experience, such as Angry Looking Jack. I argued that bare appeals to intuition about the key cases were controversial, not

clearly about justification, and incorrectly described by proponents of the challenge.

The argument from emotion does significantly better here. It does not rely solely on controversial intuitions about the key cases of cognitively penetrated visual experience. In fact, the argument from emotion does not require highly controversial appeals to intuitions about cases of vision or emotion. Instead, the argument employs widely-accepted claims in the psychology and philosophy of emotion, as well as widely-accepted claims in epistemology, as support for its premises. I provided evidence that our assessments of the emotional states and their etiologies concern a form of epistemic justification. And the points I made hold even if we take into account the points I raised about introspection in §3.

5.2 The Testimony Analogy Revisited

In §3.2, I discussed an attempt to establish the key verdicts by appeal to an analogy with testimony. I argued that, once properly formulated to excluding defeating evidence, the Disguised Gossip Circle case did not support the key verdicts. Moreover, I suggested that the analogy between cognitive penetration and gossip circles is vulnerable to attack. In the gossip circle cases, the causal chain traces outside the subject's mental system, whereas the causal chain in cases of cognitive penetration traces within the subject's mental system.

The emotion analogy again does better here. First, the key verdicts applied to the cases of emotion are plausible and do not require the presence of defeaters.

Second, the etiologies involved in cognitive penetration of emotional and visual experiences are much more similar to one another than either is to a gossip circle. In particular, both cognitive penetration processes trace path's entirely within the subject's cognitive system.

5.3 The Belief Analogy Revisited

In §3.3, I discussed an attempt to establish the key verdicts by appeal to an analogy with beliefs. I argued that the putatively analogous cases of belief and visual experience do not have sufficiently similar etiologies or phenomenology.

The emotion analogy avoids these problems for the belief analogy. Regarding etiology, I extended the notion of cognitive penetration to cover both visual and emotional experiences, and I described how the psychological processes in both types of case are relevantly similar. Regarding phenomenology, I argued that the cases of visual and emotional experience have similar presentational phenomenology at least partly in virtue of which both kinds of state can provide justification when they do. The similarity contrasts with Siegel's belief examples, which do not have presentational phenomenology and may have no phenomenology at all.

6 Conclusion

In this chapter, I have offered a novel line of argument for the key claims in the new challenge from cognitive penetration for theories of perceptual justification. I

presented the argument by analogy with emotion, defended it against some objections, and explained how the argument avoids the problems raised for previous attempts to establish the key claims.

I conclude with a note about the roles of phenomenology and etiology in the justification provided by the relevant states—including both emotional and visual states—and some of the possible larger implications of the paper. In presenting the examples and arguments in the present paper, I have suggested that when visual and emotional states provided justification, they may do so partly in virtue of their distinctive phenomenology. My remarks on this score put me in agreement with views such as dogmatism that emphasize the justificatory role of phenomenology and in disagreement with views such as reliabilism which do not (e.g. Goldman, 1986). On the other hand, I have also argued that the etiology of an experience can prevent it from providing the usual justification for external world beliefs. My remarks on this score put me in agreement with reliabilism, which emphasizes the justificatory role of etiology, and in disagreement with views such as dogmatism, which gives no such role to etiology.

A larger conclusion of the paper, then, is that both the phenomenology and the etiology of an experience have distinctive roles to play in fixing the justification provided by an experience. This claim is most readily seen when looking at cases of emotional experience. But if the analogy I have presented here between emotional and visual experiences holds, we have grounds for novel support for theories of

perceptual justification that identify justificatory roles for both the phenomenology of perceptual experiences and their etiologies.

CHAPTER 2

THE COGNITIVE PENETRABILITY CHALLENGE TO RELIABILISM

0 Introduction

Skepticism aside, it should be uncontroversial that the etiology of a perceptual experience can play an indirect role in fixing the epistemic justification of beliefs based on the experience. An experience's etiology helps fix its phenomenology and content, which in turn help fix the relevant justification. However, it is controversial whether two experiences identical in their content and phenomenology, but differing in their etiology, can provide different (degrees of) justification.

Theories of perceptual justification differ in the justificatory roles they identify for the phenomenology (i.e. 'what it's like') and etiology of experiences. For example, reliabilists emphasize the role of etiology in fixing justification, while dogmatists emphasize the role of phenomenology. Recently, reliabilists including Goldman (2008a) and Lyons (2011) have endorsed an epistemic challenge from cognitive penetrability against rivals such as dogmatists. The challenge supposedly gives reliabilists an advantage over their rivals, assuming reliabilists avoid the challenge. In this paper, I argue that the challenge arises for reliabilism. Thus, the challenge affects more theories than has been recognized. Epistemologists of various stripes should investigate the conditions under which the etiology of an experience affects the justification it provides.

Chapter 2 is organized as follows. In §1, I introduce reliabilism. In §2, I review what cognitive penetrability is and present an example of it not covered in Chapter 1. In §3, I explain how the challenge arises for previous targets. In §4, I argue that the challenge arises for reliabilism. In §5, I consider a response on reliabilists' behalf and rebut it. In §6, I show how two familiar problems for reliabilism—the generality and new evil demon problems—bolster the cognitive penetrability challenge. I conclude in §7.

1 Reliabilism

Reliabilism comes in various forms. I begin by explaining the forms I target. *Pure reliabilism* entails that a belief is doxastically justified *iff* it is produced or sustained by a reliable, relevant process-type.[38] Because token processes are not assessable for reliability, reliabilism applies to process-types. There are indefinitely many types under which each token process falls, and the types' reliability varies. For reliabilism to yield determinate verdicts about justification, there must be a unique, relevant process-type whose reliability fixes the degree of justification for the belief for each instance (or, alternatively, ascription) of justification.

Many reliabilists reject pure reliabilism in response to examples like BonJour's (1980) clairvoyant Norman and Lehrer's (1990) Mr. Truetemp. Those examples—especially BonJour's—are too fantastical for some readers. More realistic examples

[38] Hereafter, I omit 'or sustained' for ease of presentation.

raise the same problems for pure reliabilism. Consider:

> **Graphology** On the basis of a person's handwriting, S can reliably form beliefs about the person's emotional state at the time of writing. Without noticing, S is sensitive to subtle graphological cues and can reliably form beliefs on the basis of these cues. S possesses evidence that no such ability exists and that she does not have the ability. One day, S reads a note handwritten by X, time stamped just moments earlier. Deploying the reliable belief formation process involving her graphological skills, S forms the true belief that X is angry.

S's belief that X is angry is doxastically unjustified. That the relevant process-type that produced S's belief is reliable is insufficient to justify the belief.

Responding to such examples, many reliabilists adopt forms of *impure* reliabilism (e.g. Goldman, 1986: 63, 111-112). According to impure forms, reliability is necessary but not sufficient for justification. Some impure forms of reliabilism add a 'non-undermining' or 'no-defeaters' condition. Here is a representative formulation.

> **Reliabilism** S's belief is doxastically justified *iff* (i) the relevant process-type is reliable and (ii) S lacks (sufficient) evidence for the claim that the relevant process-type is not reliable.[39]

[39] This simplified formulation represents a family of views that go under the label 'reliabilism'. It is derived from Goldman (2008b). I discuss additional formulations of condition (ii) in §4.2. There are a number of formulations of impure reliabilism about justification that I will not discuss in the main text but to which my arguments apply with equal effectiveness. The examples I present also raise problems for the two-stage attribution approach adopted by Goldman

The no-defeaters condition is supposed to immunize Reliabilism from counterexamples like Graphology. S has evidence that graphology-based belief-formation processes are unreliable. So she has a defeater for justification she might have had from the reliable graphological process to believe that X is angry. The no-defeaters condition is not an *ad hoc* response to counterexamples, however. Justification is defeasible, and a no-defeaters condition is required to account for defeasibility.

In what follows, I argue that the cognitive penetrability challenge arises for both pure and impure forms of reliabilism. However, like others, I take pure reliabilism to be refuted by counterexamples like Graphology. My primary aim, then, is to raise the cognitive penetrability challenge for impure reliabilism.

2 Cognitive Penetrability

As noted in Chapter 1, 'cognitive penetrability' can be defined for various kinds of states. In this chapter, I focus on visual states. Recall that in Chapter 1 'cognitive penetrability' was defined for visual states as follows: a visual experience is cognitively

---

(1992), since one attributes an unjustified belief to the subjects in the relevant examples even after learning that the process-types in question are reliable. The challenge also arises for reliabilism on a number of versions articulated in response to the new evil demon problem. For example, because all of the examples I use in the paper occur in the actual world or nearby possible-worlds, the challenge arises equally for versions of reliabilism using a 'normal worlds' approach (Goldman, 1986: 107). In addition, a number of versions of reliabilism that differ in their solution to the generality problem face the challenge, such as Heller (1996) and Comesana (2006). Finally, I should reiterate that the challenge

penetrable with respect to some content or aspect of phenomenal character *c* if and only if two subjects (or the same subject at different times) can differ with respect to whether their experience has *c*, and the difference is the result of a causal process tracing back to a non-visual, psychological state of the subject, where we hold fixed between the two subjects (or one subject at different times) the following: (i) the stimuli impacting their sensory receptors, (ii) the subjects' spatial attention, and (iii) the conditions of the subjects' sensory organs.[40]

Two examples of cognitive penetration illustrate the phenomenon and feature in my discussion of the cognitive penetrability challenge to Reliabilism. The first was featured in Chapter 1; it involves a penetrating belief.

> **Angry Looking Jack** Before seeing Jack, Jill has the true but unjustified belief that Jack is angry. When she sees him, Jill's unjustified background belief causes her to have a visual experience in which Jack looks angry. If she had lacked the belief, Jack would not have looked angry to her. In addition, Jill is reasonably ignorant of the causal role her background belief plays. As the result of her experience, Jill reaffirms her belief that Jack is angry. (Siegel, 2012: 2)

In Angry Looking Jack (ALJ), Jill's background belief that Jack is angry causes her to

---

arises for reliabilist theories of justification; I do not argue that it arises for reliabilist accounts of knowledge such as Armstrong (1973) and Sosa (2007).

[40] For additional discussion, see Ch. 1, note 4.

have a visual experience as of Jack's being angry. The content and phenomenology of Jill's experience would have been different had she not held the background belief. The difference would obtain even if we held fixed the stimuli impacting Jill's sensory organs and her spatial attention. As a result, ALJ involves cognitively penetrated visual experience.

The second example involves a penetrating desire.

> **Wishful Grade** Sam desires a B in History. Unfortunately, he got a D. When he looks at his report card, his desire causes it to look to him as though the grade is a B. If he had lacked the desire, it would have correctly appeared as a D. Sam is reasonably unaware that his background desire influences his experience. Sam believes that he got a B in History based on his penetrated experience.[41]

Sam's desire influences his experience via causal paths within Sam's mental system and affects his experience without requiring a change in spatial attention or distal stimuli. So Wishful Grade is a case of cognitive penetration.

[41] Wishful Grade is derived from a similar example in Markie (2005) involving two prospectors looking for gold. For both subjects, it seems to them that a nugget is gold. For one prospector, the seeming is due to his desire to find gold, for the other it is due to his learned identification skills. Markie's main target with the challenge is Huemer's (2001) phenomenal conservatism, which is supposed to apply to any 'seeming' state, not just visual states. Markie's case is not underdescribed for that purpose, but for my purposes it is. It is not clear from Markie's description what kind of phenomenology the two prospectors enjoy, especially whether it is visual or not. Wishful Grade makes it clear that the phenomenological effects of the desire penetrate to visual experience.

3 Previous Targets of the Challenge

As noted in Chapter 1, a number of recent authors have used examples like Wishful Grade and ALJ to raise a challenge for various theories of perceptual justification.[42] For my purposes in this chapter, the key epistemic verdict about the cases is as follows:

Key Verdict The subject's post-experience belief is doxastically unjustified.

In what follows, I assume that the key verdict is true. I have already argued for it in Chapter 1, when arguing for verdict (b). In addition, as part of my challenge to Reliabilism, taking the truth of the key verdict for granted is dialectically fairly secure. Goldman argues that cases like Wishful Grade pose a challenge to dogmatism (Goldman, 2008a: 73). Lyons (2011) argues that ALJ does. Finally, the cognitive penetrability challenge does not get off the ground—and thus cannot provide an advantage for reliabilists—unless the key verdict is true.

Views for which the challenge arises are supposed to be inconsistent with—or at least in tension with—the key verdict. Most discussion has been directed toward versions of dogmatism as a target of the challenge. Recall the view:

[42] In addition to Goldman (2008a) and Lyons (2011), proponents of the challenge include Markie (2005, 2006), Siegel (2011, forthcoming-a), Jackson (2011), and McGrath (forthcoming-a, forthcoming-b).

Dogmatism A visual experience as of *p* provides some defeasible, immediate propositional justification for *p* in virtue of its having a distinctive phenomenology with respect to *p*.

Recall that ALJ, for example, is supposed to raise a challenge for dogmatism as follows. Jill enjoys an experience with the right phenomenology to provide justification for the claim that Jack is angry, according the dogmatism. Jill bases her reaffirmed belief on her experience. Further, she is reasonably unaware of the cognitive penetration process, so she arguably lacks defeaters for any justification the experience might provide. Dogmatists are thus under pressure to accept the claim that Jill's post-experience belief is doxastically justified, contrary to the key verdict. At the very least, dogmatism seems ill-positioned to account for the truth of the key verdict. A full answer to the cognitive penetrability challenge not only shows that one's theory of (perceptual) justification is consistent with the key verdict; one must identify the features of cases of cognitive penetrability that account for the key verdict.

Reliabilists at first appear better positioned than dogmatists to answer the challenge. According to Reliabilists, facts about the etiology of an experience (and the resulting beliefs) play a central role in fixing perceptual justification, facts beyond those concerning the etiology's role in fixing the content and phenomenology of the relevant experience. Moreover, it may initially appear likely that Sam's and Jill's beliefs are produced by unreliable process-types. So, Reliabilism at first appears consistent

with the key verdict and well-positioned to account for its truth. However, while reliabilists might be right to give the etiology of experiences an important role in their account of justification, there is no guarantee that they have identified all, or indeed, any of the etiological features that help fix the relevant justification.

4 The Challenge to Reliabilism

In this section, I argue Reliabilism is inconsistent with the key verdict by arguing that Reliabilism's conditions (i) and (ii) are satisfied in the cases. Establishing that condition (i) is satisfied is sufficient for a counterexample to pure reliabilism. Establishing that both are satisfied is sufficient for a counterexample to the version of impure reliabilism above.

4.1 On Condition (i): Reliability

4.1.1 Candidate Process-types

In this sub-section, I survey a range of candidate process-types and argue that each is reliable.[43]

First, perhaps the relevant process-type is *the process-type that produces perceptual*

[43] There are at least two kinds of reliability one might appeal to here: categorical and conditional reliability. A process-type is categorically reliable *iff* it tends to produce a high proportion of true over false beliefs. A process-type is conditionally reliable *iff* it tends to produce a high proportion of true over false beliefs *given that all its input-beliefs are true* (Goldman, 1979). In the main text, I will treat the discussion as though it is in terms of categorical reliability. However, it should be noted that appealing to conditional reliability will not help reliabilists respond to the examples. The same arguments used to establish the categorical reliability of the candidate process-types are equally effective in establishing that these process-types are conditionally reliable. In ALJ, Jill's pre-experience belief that Jack is angry is true, and we may suppose that all of the input-beliefs in the cognitive penetration process that yields Jill's post-experience belief are

beliefs. The type is a common-sense kind and arguably a scientific (natural) kind—two plausible criteria for identifying the relevant process-type.[44] Jill's case falls under the type. However, perception is reliable. If this is the relevant type, it does not help reliabilists avoid the challenge. A similar suggestion—that the relevant type is *the process-type that produces beliefs on the (immediate) basis of visual experience*—faces the same fate. Visual processing is reliable.

A more promising suggestion is that the relevant type is *the process-type that produces beliefs on the basis of cognitively penetrated visual experience*. The suggestion provides a good candidate for the relevant type. One might think that for visual processing to reliably produce true beliefs it must typically represent one's environment without distortion. Cognitive penetration seems to have a distorting effect on vision, and one might conclude that cognitively penetrated visual processing is an unreliable process-type.

To assess the proposal we need evidence for the unreliability of beliefs formed on the basis of cognitively penetrated visual experience. Initially, the proposal appears to find support from a famous experiment by New Look psychologists Bruner and Goodman (1947). Children were divided into two groups—'rich' and 'poor'—based on their socio-economic background. The children viewed coins of varying denominations and were asked to adjust a small light patch to match the size of each

---

true, as well. With these stipulations, Reliabiliism entails the same verdicts for the case whether the subjects' beliefs are assessed for categorical or conditional reliability.

[44] For these proposals, see Conee and Feldman (1998).

coin. Results suggested that both sets of children adjusted the light in overestimation of the coins' actual size (and relative to their adjustments for coin-sized paper discs). Additionally, 'poor' children adjusted the light patch in overestimation of the coins' size by up to 50% more than 'rich' children. The researchers hypothesized that 'poor' children had a greater desire for the money which penetrated their visual experience of the coins' size, causing them to appear larger.

However, Bruner and Goodman's results were not consistently replicated. Problems with the original experimental design were uncovered. For example, the adjustable light-patch used by Bruner and Goodman wasn't filled in completely and lacked sharp boundaries like coins. When Carter and Schooler (1949) used a solid, sharp-edged light patch, subjects' overestimation was greatly reduced. In addition, Carter and Schooler compared size estimates of coins and valueless metal discs of corresponding size; they failed to find statistically significant overestimations of the coins compared to the discs. In response to these and other criticisms, Bruner and Rodrigues (1953) ran a series of more carefully controlled, better-designed follow-up studies. They too failed to replicate the more striking earlier results (Bruner & Rodrigues, 1953).[45] For example, they too failed to find statistically significant differences between subjects' size estimations of coins and similar-sized, valueless metal discs.[46]

[45] See especially Bruner and Goodman's Table 2.

[46] They write, "As in the Bruner-Goodman study, there is a significant difference between coins and cardboard discs… But it is quite apparent that in terms of absolute level of accentuation, our results are negative where a comparison of

Bruner and Rodrigues did obtain some statistically significant results, however. The results suggest that cognitive penetration may actually increase the reliability of subjects' judgments. While subjects tended to make fairly accurate estimations for pennies, they made increasingly inaccurate estimations of the sizes of larger coins, accentuating the differences in their size judgments between the denominations. In a good review of this early literature, Tajfel (1957) argues that the results across all studies point to the following conclusion: subjects' knowledge of the correlation between a coin's size and its denomination penetrates their experience, affecting the apparent size, thereby helping subjects reliably identify coins by denomination. Since many of our beliefs about coins concern their denominations and not precise size, the trend in the results suggests that cognitive penetration may enhance the reliability of our judgments.

Additional evidence of the reliability-enhancing power of cognitive penetration comes from recent work on color vision. In a series of recent studies, Thorsten Hansen, Karl Gegenfurtner, and colleagues have found evidence that memory modulates color experience. Their results suggest that memory accentuates the characteristic colors of familiar objects. For example, Hansen et al. (2006) found that subjects perceived images of bananas as yellower than images of unfamiliar objects with the same surface-reflectance properties. They found similar results for a variety

---

coins and metal discs is concerned. To be sure, an analysis of variance, which we shall discuss more freely below, shows that type of object being judged is a significant source of variance, but this result is a function of something other than any consistency in overestimation of the coins relative to the metal discs. It probably reflects the greater apparent

of other fruits with characteristic colors. Olkkonen et al. (2008) replicated the results for fruits under a variety of viewing conditions. They also found that the accentuation is most pronounced for the most detailed, realistic fruit images, as opposed to vague, unrealistic outlines of the fruit. The results suggest that the relevant process-type shaping color perception typically accentuates the characteristic color of real fruit objects without being misled by similar objects. Finally, Witzel et al. (2011) achieved similar results for realistic images of human-made artifacts with familiar, characteristic colors such as blue Smurfs and red Coca-Cola logos. These additional findings suggest that the modulating effects sometimes result from subjects' acquired knowledge of objects' characteristic color. The process plausibly qualifies as cognitive penetration on the definition provided above.

Taken in total, the color modulation results suggest that the cognitive penetration of color vision by acquired knowledge of characteristic color helps improve the accuracy of color representation across a range of viewing conditions. The data also suggest that cognitive penetration of color vision boosts the reliability of perceptual belief formation by improving subjects' ability to recognize objects of importance to them by accentuating the characteristic colors on the basis of which they can be recognized.

It is worth remembering that reliabilists are concerned with the reliability of belief-forming (and sustaining) processes. They are not concerned in the first instance

---

magnitude of the two metals as compared with the paper" (Bruner & Rodrigues, 1953: 20).

with the accuracy of processes except insofar as those processes produce or sustain beliefs. The point is important, since the evidence from the studies cited above suggests that cognitive penetration may improve the reliability of processes leading to beliefs (e.g. about the type of object seen) at the expense of the accuracy of processes that typically do not lead to beliefs (e.g. about the precise size or color of objects). Cognitive penetration may improve the reliability of belief-forming processes by reducing the reliability of other non-belief-forming processes.

Nevertheless, penetrating states can mislead subjects and cause them to form false beliefs. One might wonder whether the distortions introduced by cognitive penetration are frequent and pronounced enough to substantially reduce the reliability of the relevant process-type. Here it is also important to remember that, even if some instances of cognitive penetration reduce reliability, not just any reduction is sufficient for the reliabilist's purposes in response to the cognitive penetrability challenge. Recall that the key verdict is that the subjects' beliefs in the examples are doxastically unjustified. Showing that cognitive penetration reduces reliability would only show that that the subjects' beliefs are less justified according to Reliabilism than subjects with qualitatively identical non-penetrated visual experiences. The reduction in reliability might not be enough to support the key verdict.

Recent empirical work on cognitive penetration of vision by desire provides further evidence that cognitively penetrated visual experience is too reliable for Reliabilism to yield the key verdict in the relevant cases. For example, Balcetis and

Dunning (2010) conducted a series of studies to measure the effects of desire on perception of distance. They found that subjects regularly underestimated the distance to desired objects—both in comparison to the actual distance and in comparison to distance estimates by subjects in relation to non-desired objects. The researchers hypothesized that subjects want desirable objects to be closer to them than they actually are, the desire penetrates their visual experience, and, as a result, subjects visually experience desired objects as being closer than equidistant non-desired objects. In one study, thirsty subjects estimated the distance to a water bottle as 10.4% closer compared with estimates by quenched subjects (Study 1). Other subjects judged a $100 bill as 13.8% closer when they had a chance of winning it than when they did not (Study 2a). Balcetis and Dunning took these results as evidence of cognitive penetration of visual experience by desire.

One might worry that the subjects' distance estimates are the result of an influence by desire on subjects' post-experience judgments without cognitive penetration of their visual experience. In an effort to control for that possibility, Balcetis and Dunning conducted a study asking subjects to stand a prescribed distance from an object. Subjects stood across from a wall with two vertical strips of tape 90.5 in. apart. On a table in front of the subjects, the researchers placed either colorfully wrapped chocolates (the desired object) or what the subjects were told was a bag of recently collected dog feces (the non-desired object). Subjects were to place themselves so that the distance between them and the object (chocolate or feces) was

equal to the distance between the tape strips. Results showed that subjects stood farther from the chocolate (mean = 101.3 in.) than from the feces (mean = 88.0 in.). The results corroborate the earlier studies' results, by suggesting that subjects perceived the chocolates as closer by comparison with the feces.

Balcetis and Dunning's results suggest that distorting effects of desire did reduce the accuracy of any precise judgments subjects made about the distances involved. The differences in distance perception between desire-influenced subjects and controls ranged from 6.7% to 16.2% across the studies. The absolute inaccuracy of desire-influenced subjects' perceptions also fell within that range. Since the effect sizes of inaccuracy were small, beliefs about the approximate distances of objects that subjects formed on the basis of such experiences may well have been as reliable for desire-influenced and other subjects. Moreover, it is highly implausible that non-desire-influenced subjects' distance judgments were perfectly accurate. Given the generally very high reliability of visual processing, the small reductions in the absolute accuracy of desire-penetrated visual experiences do not reduce the reliability of distance judgments by much, if at all. Thus the results suggest that the distorting influences of cognitive penetration are insufficient for the relevant process-type to be unreliable enough for Reliabilism to yield the key verdict.[47]

[47] Additional evidence for this conclusion comes from the color vision studies mentioned above, where researchers found similarly small effect sizes. To measure the effects precisely, Hansen et al. defined a Memory Color Index (MCI) which specified the degree of shift between a subject's white point and full saturation for the relevant color. Using the MCI to measure the surface reflectance properties of the images that subjects reported experiencing as achromatic, the researchers were able to specify the degree of shift. They found only 3% to 14% shifts in the MCI of the experience of

To close the empirical discussion of cognitive penetration's distorting effects, it is worth noting that many plausible cases of cognitive penetration do not involve distortion. For example, an adult's knowledge of pine trees might penetrate their visual experience causing it to represent the content that something is a pine tree, whereas, the visual experience of a naïve subject without knowledge of pine trees would not represent that content. Or a doctor's visual experience might represent that a patch on an x-ray is a tumor, whereas the experience of a subject lacking the relevant medical training would not represent that content.[48] These differences in the contents a visual experience represents can result without shifts in attention or distal stimuli; that is, they can result from cognitive penetration. Such cases may commonly occur, and they need not involve distortion. The common occurrence of such cases provides further evidence for the reliability of the process-type that produces beliefs on the basis of cognitively penetrated experience.

To conclude this sub-section, I consider one additional process-type reliabilists could appeal to in response to the challenge. It is the process-type suggested by Goldman (2008a) in his discussion of the challenge. Goldman focuses on a case of desire-influenced cognitive penetration like Wishful Grade and suggests *wishful thinking* as the relevant process-type. We can grant that wishful thinking is an unreliable

different fruits (mean = 8.23%, $p < .001$) as compared with the subjects' experience of objects with the same surface reflectance properties but no characteristic color. The MCI numbers represent shifts at only about 3 to 5 times above the threshold of discrimination (Hansen et al., 2006: 2). Given the large number of discriminable hues, shades, and the like, these are very small shifts. So, for example, although subjects experience bananas as yellower than unfamiliar objects with the same surface reflectance properties, they only experienced the bananas as slightly yellower than those other objects. Olkkonen et al. and Witzel et al. found similar degrees of shift.

process-type. It is also a good candidate for the relevant process-type for Sam's post-experience belief in Wishful Grade. Even so, the proposal fails to address all of the relevant cases. ALJ is clearly not a case of wishful thinking and must still be dealt with. As the preceding discussion suggests, that is no easy task for reliabilists.

In fact, Goldman's proposal may make the task of responding to examples like ALJ more difficult for reliabilists. Wishful thinking is a fairly general process-type. Appealing to it to handle Wishful Grade puts reliabilists under pressure to appeal to a similarly general process-type to handle ALJ and other cases. But, given the empirical data surveyed above, it is unlikely that there is a general type that will fill the bill. For example, as I argued above, perception, vision, and cognitive penetration processes are all reliable process-types. Rather, it seems that the kind of process-type that explains our intuitions about Jill's case would have to be quite specific. Thus, even if appealing to wishful thinking handles Wishful Grade, it may cause more problems for reliabilsits than it solves.

4.1.2 Filling in the Counterexample

So far I have pressed the challenge for pure and impure reliabilism alike by surveying a number of candidates for the relevant process-type and arguing that each is reliable. I now explain how the challenge may arise differently for the two forms of reliabilism. Recall that pure reliabilism entails that the subjects' beliefs are justified if and only if

[48] See Siegel (2006) for arguments that visual experience can represent high-level properties such as *being a pine tree*.

they are produced by a reliable, relevant process-type. Pure reliabilists might respond to the two main examples I have discussed here in the following ways. Regarding Wishful Grade, they might identify an unreliable process-type (e.g. wishful thinking). Regarding ALJ, they might first note that, to secure the stipulation that Jill's pre-experience belief is unjustified, that belief must be produced by a relevant, unreliable process-type. Pure reliabilists could then argue that the relevant process-type for Jill's *post*-experience belief that Jack is angry includes the process-type that led to her *pre*-experience belief as a proper part. Finally, they could argue that the whole relevant process-type, including the unreliable part leading to Jill's pre-experience belief, is unreliable. They need not argue that it is always the case that a process-type with an unreliable proper part is unreliable, just that it is in Jill's case.

It is not obvious that the relevant etiology includes the process-types that led to Jill's pre-experience belief, but it is plausible that it does. So the response is promising for pure reliabilists. However, as I noted at the outset, I think that pure reliabilism faces other, familiar counterexamples. So my main concern in the present paper is to elaborate a novel challenge from cognitive penetrability for impure forms of reliabilism.

Impure reliabilists cannot answer the challenge posed by ALJ in the way pure reliabilists can. We can fill in the backstory of Jill's pre-experience belief so that it is produced by a reliable process-type yet unjustified, according to impure reliabilists (but not according to pure reliabilists). The elaborated example has two stages. The

additional backstory in Stage 1 utilizes the details from the example (that motivated the move from pure to impure reliabilism. Stage 2 reproduces the details from ALJ.

> **Graphological ALJ** *Stage 1* (Graphology) On the basis of a person's handwriting, Jill can reliably form beliefs about the person's emotional state at the time of writing. Without noticing, Jill is sensitive to subtle graphological cues and can reliably form beliefs on the basis of these cues. Jill possesses evidence that no such ability exists and that she does not have the ability. One day, Jill reads a note handwritten by Jack, time stamped just moments earlier. Deploying the reliable belief formation process involving her graphological skills, Jill forms the true belief that Jack is angry. *Stage 2* (ALJ) When Jill sees Jack moments later her background belief that Jack is angry causes Jill to have a visual experience in which Jack looks angry. If she had lacked the belief, Jack would not have looked angry to her. In addition, Jill is reasonably ignorant of the causal role her background belief plays. As the result of her experience, Jill reaffirms her belief that Jack is angry.

Graphological ALJ differs from the original ALJ by adding a backstory for Jill's pre-experience unjustified belief. With only this addition, the key verdict that Jill's post-experience belief is doxastically unjustified remains compelling. However, (impure) Reliabilism predicts a contrary verdict. Jill's pre-experience belief satisfies Reliabilism's

reliability condition (i), but it fails the no-defeaters condition (ii). So Reliabilism entails that Jill's pre-experience belief is unjustified. As noted, the process-type that leads to Jill's pre-experience belief is reliable. And, as I argued in the previous sub-sections, the process-type leading from Jill's pre-experience belief to her post-experience belief is also reliable for a range of candidates. So even if we include the causal processes leading to Jill's pre-experience belief in etiology of the post-experience belief, the relevant process-type is a reliable one. Condition (i) is satisfied for the case. As before, Jill is reasonably unaware of the cognitive penetration process, so she plausibly lacks defeating evidence for her *post*-experience belief's justification (more on this in the next sub-section). So Reliabilism falsely entails that Jill's post-experience belief is justified.

In this sub-section, I have surveyed a range of candidates for the relevant process-type in Wishful Grade, ALJ, and Graphological ALJ. I have argued that each candidate is reliable (perception, vision, and cognitively penetrated vision) or leads to more problems for reliabilists than it solves (wishful thinking). I conclude that none of the candidate process-types answers the challenge.

4.2 On Condition (ii): Defeat

In this section, I argue that condition (ii) is satisfied in Wishful Grade, ALJ, and Graphological ALJ. I offer two arguments that (reliabilists are committed to the claim that) Jill lacks defeaters for her post-experience belief that Jack is angry. I then

consider modifications reliabilists could make to (ii). Recall condition (ii):

(ii) S lacks (sufficient) evidence for the claim that the relevant process-type is not reliable.

The first argument that condition (ii) is satisfied is as follows: in the relevant cases, the subject lacks evidence that cognitive penetration is operative in their case. As far as they can tell, they experience normal vision. They have excellent reason to believe that their post-experience belief is the result of ordinary visual experience, and they have excellent reason to believe that ordinary visual experience is reliable. So the relevant subjects do not possess a defeater whatever justification may be provided by their experience for their post-expereince belief.

The second argument is a *tu quoque*. Reliabilists argue that the cognitive penetrability challenge arises for views like dogmatism. The challenge arises only if subjects in cases like Wishful Grade and Graphological ALJ lack defeating evidence for their post-experience beliefs. If they possessed defeaters, dogmatists could point to the presence of defeaters to explain the key verdict, and dogmatism would not entail the negation of the key verdict. Reliabilists endorsing the challenge must claim that condition (ii) is satisfied.

Reliabilists could modify condition (ii). Perhaps reliabilists should not build in a condition that says a reliably produced belief is justified unless the subject *possesses*

evidence of the unreliability of the process-type. The modified condition might then be the following:

(ii)* S lacks defeaters for her justification for *p*.

(ii)* does not say that the relevant subjects lack evidence for the unreliability of the relevant process-type. It is in principle compatible with the claims that the relevant subjects possess defeaters. However, it does not solve the challenge. If reliabilists appeal to (ii)* to respond to the challenge, they must say more about how to interpret it. Either the condition employs the same conception of (the conditions of) defeat as features in the version of the challenge this is supposed to arise for dogmatism, or it does not. Neither option is promising for reliabilists as a solution to the challenge.

If (ii)* employs the conception of defeat that was supposed to be common ground between reliabilists and dogmatists in previous discussions of the challenge, it is satisfied in Wishful Grade, ALJ, and Graphological ALJ. Above I noted that for the challenge to get off the ground against dogmatists, Jill must lack defeaters for any justification she receives from her cognitively penetrated experience. But, if Jill lacks defeaters, (ii)* is satisfied, the modified form of reliabilism still falsely entails that Jill's post-experience belief is justified. So, if (ii)* employs the conception of defeat that was supposed to be common ground between reliabilists and dogmatists in previous discussions of the challenge, the proposal fails to answer the challenge.

In order to get a no-defeaters condition that will help reliabilists respond to the challenge, the modified condition must employ a different conception of (the conditions of) defeat other than the one that is employed in raising the challenge for dogmatists. There are several problems with the approach. First, it is not an effective strategy to argue for the advantages of Reliabilism over dogmatism, for example, unless reliabilists argue that they can—but dogmatists cannot—appeal to the relevant claims about (the conditions of) defeat. It is not clear that such an argument is available.[49]

Second, it is not clear what the modified conception of (the conditions of) defeat should be. It will not do to conceive of a defeater as *anything that prevents justification in the circumstances*. On such a conception, the modified form of reliabilism would be no more informative than the following view: a belief is justified *iff* it is produced by a reliable, relevant process-type and nothing prevents the belief from being justified. The view is trivial and leaves no work for reliability in the account. Moreover, to answer the challenge, one cannot simply say that the examples reveal the need for a modified account of defeat without spelling out such an account. Appealing to defeaters merely pushes the problem back. The hard work of identifying

[49] One might think that there is a reason reliabilists, but not dogmatists, can adopt such an altered conception of defeat. Dogmatists tend to adopt one or another form of access internalism about justification, while reliabilists tend to be externalists. Perhaps the needed claims about (the conditions of) defeat will require an externalist approach well-suited to Reliabilism but not dogmatism. However, accepting dogmatism does not require one to accept access internalism about justification. And, as noted above, dogmatism is not strictly a view about doxastic justification. It is a view about defeasible (*prima facie*) justification. Dogmatism entails nothing about the nature or conditions of defeat. Thus it is hard

the defects in the relevant cases remains.[50]

5 Belief-Dependent Justification

In the previous section, I argued that the cognitive penetrability challenge arises for Reliabilism. In the remainder of the paper, I further press the challenge by providing additional arguments that answering the challenge is no easy task for reliabilists. In this section, I consider further possible responses reliabilists might make, and I raise worries for them. In §6, I show how two other challenges to Reliabilism compound the difficulties raised by cognitive penetrability.

Perhaps Reliabilism should include a condition for justification dependent on the justification-statuses of other beliefs. Some cases of cognitive penetration have beliefs are inputs; for example ALJ. That is, some instances of cognitive penetration are belief-dependent processes. Reliabilists could propose assessing output beliefs' justification depending on the input-beliefs' justification. Compare: deductive inference is a belief-dependent process, and the justification-statuses of the deductively inferred beliefs arguably depend on the justification-statuses of premise-beliefs. For example, suppose that S infers *q* from a justified belief that *p* and an

---

to see what would prevent dogmatists from adopting the claims that reliabilists might appeal to in modifying condition (ii), other than dogmatists' independent commitments concerning the nature and conditions of defeat.

[50] More generally, the concept of a defeater is probably not sufficiently well-understood to be used as leverage to make the challenge tractable, especially if reliabilists must introduce modified characterizations of the concept. John Hawthorne and Maria Lasonen-Aarnio, for example, argue that, in epistemology, the concept of a defeater is "woefully underdeveloped and overdeployed" (Hawthorne & Lasonen-Aarnio, 2009). While this may be an overstatement of the problem for appeals to defeat in the literature, there does seem to be a lack of clarity about the conditions under which

unjustified belief that (If *p, q*). S's belief that *q* is not justified, assuming that S lacks some independent source of justification for *q*.

The proposal fails to address the challenge posed by cases like Wishful Grade, where the improper etiology does not include an input-belief, justified or unjustified. So, at best, it could serve as an incomplete response to the challenge. In addition, the proposal requires modifying Reliabilism to include an additional condition concerning belief-dependent justification. It is not clear whether such a condition falls within the spirit of Reliabilism. Still, the proposal is worth investigating. To test the proposal, I consider several possible formulations of the condition. I argue that none of them succeeds in answering the challenge.

One possible formulation is as follows: for an output-belief *b* that results from a conditionally relia ble belief-dependent process, *b* is justified only if all of the input-beliefs to the process are justified.[51] Applying the proposal to Graphological ALJ yields the following: Jill's post-experience belief is produced by a belief-dependent process of cognitive penetration with one unjustified input-belief. So the proposal entails that her output-belief is unjustified, the true verdict.

---

and in virtue of which justification is defeated. One way to cast the cognitive penetrability challenge connects it with the need for additional clarity concerning the nature of defeat and the conditions under which justification is defeated.

[51] See Lyons (2011) for this formulation as one option for how one could respond to the cognitive penetrability challenge. Lyons rejects this line of response on behalf of targets of the challenge. He notes that the proposal fails to fully address the problem, since not all instances of improper cognitive penetration involve unjustified input-beliefs. He also argues that there are some cases of cognitive penetration involving unjustified input-beliefs do not involve epistemically improper etiologies. However, Lyon's arguments here are not compelling. On the first point, even if the proposal fails to address all the cases, it might be part of a successful response. On the second point, Lyons's examples of justificatory experiences caused by unjustified background belief involve shifts in spatial attention and distal stimuli, so they do not count as cases of cognitive penetration on the definition used here. For an earlier reliabilist discussion of belief-dependent justification, though not in the context of the cognitive penetrability challenge, see Goldman (1979).

The above formulation fails to answer the challenge. Even if it can be motivated as a way of handling the Graphological ALJ, it leads to further counterexamples, such as the following.

> **Subtle ALJ** Same as 'Graphological ALJ', except: The subject is Jill*, and the effect of cognitive penetration of the unjustified background belief on the phenomenal character of Jill*'s experience is very subtle. Unlike in Graphological ALJ, if Jill* hadn't held her unjustified background belief that Jack was angry prior to her experience as of Jack's face, Jack's face would have looked angry. The input-belief causes only subtle shift from one paradigmatically angry look to a slightly different angry look.

Unlike Jill's belief—which was intuitively unjustified—Jill*'s post-experience belief is intuitively *justified*, despite the fact that her unjustified background belief affects the precise phenomenology of her experience. In Jill*'s case, the effect of the unjustified background belief is too small—it can be arbitrarily small—to have such a large impact on the status of the output-belief. However, on the current proposal, Reliabilism entails that Jill*'s output-belief is unjustified, since there is at least one unjustified input-belief in the process that yields it. The current proposal fails because it makes the existence of one unjustified penetrating belief sufficient to make the output-belief unjustified. Some input-beliefs play too insignificant a role to have such

a large influence on the justification status of the output-belief.

In response, reliabilists could offer an alternative formulation of the condition, specifying that only input-beliefs playing an *essential* role in causing the output-belief help fix the output-belief's (conditional) justification. We can motivate the suggestion using a further comparison with deduction. Suppose that S infers *q* from justified beliefs that *p* and (If *p, q*) and that S also infers *q* from unjustified beliefs *r* and (If *r, q*). Since the latter beliefs are inessential to S's coming to believe *q*, S might still be justified in believing *q* on the basis of her valid inference from the justified beliefs that *p* and (If *p, q*). In similar fashion, reliabilists could argue that the justification of Jill*'s post-experience belief that Jack is angry does not depend on her pre-experience penetrating belief's justification-status, because the pre-experience penetrating belief is inessential.

To carry out the above line of response, reliabilists must clarify what 'essential' means, for it is not immediately clear how to differentiate essential from inessential beliefs in this context. However, on several natural clarifications of the notion, the suggestion fails to answer the challenge. For example, in light of the comparison with deductive inference, one could propose that an input-belief is essential only if it is *logically* essential to the validity of the (quasi)inferential process leading from the input-beliefs to the output-belief. However, there are at least two problems for differentiating essential from inessential beliefs this way. First, the logical relationships between input-beliefs and output-beliefs in a process of cognitive penetration need

not be deductively valid to be acceptable. The differentiation is too strict. To remedy the problem, one could try to relax the condition to cover a wider range of relationships including, for example, involving probabilistic coherence between input and output-beliefs. However, a second problem arises for both the strict and relaxed versions of the proposal: it is not at all clear that the processes involved in the cognitive penetration of visual experience are inferences, 'quasi-inferences' or can be modeled on (quasi)inferential processes.[52]

Taking up a different possibility, perhaps input-beliefs are essential *iff* they play an essential *causal* role in producing the output-belief. To assess the proposal, we can use a counterfactual test for essential causal contribution, applied to an example of cognitive penetration.[53] Here is the test: if S would have believed *p* whether or not she believed *q, q* is (causally) inessential to S's belief that *p*. We can see, however, that reliabilists still face counterexamples on this proposal. Consider the original ALJ case, with another added stipulation:

> **Insensitive ALJ** Same as the Graphological ALJ (where the intuitive verdict is that Jill's post-experience belief is unjustified), with the following addition: Suppose that Jill would have affirmed that Jack is angry on the basis of a non-penetrated experience as of a non-angry face she would have had, had she not

[52] For the suggestion that they can be so modeled, see McGrath (forthcoming-a). For criticism of this approach, see Siegel (forthcoming-b)

held the pre-experience belief that Jack is angry.[54]

The added stipulation—that Jill would have believed Jack was angry even without her cognitively penetrated experience—does not undermine the verdict that Jill's post-experience belief is unjustified. If anything, the stipulation reinforces the verdict. However, with the added stipulation, Jill's unjustified background belief is causally inessential to the formation of her belief that Jack's angry: as specified in the example, she would have believed it anyway. On the current proposal, Reliabilism falsely entails that (the original) Jill's post-experience belief is justified.

It should be clear that proposals adding a condition concerning essentialness in *justificatory* terms are non-starters. The challenge requires specifying the conditions under which background beliefs play their different justification-related roles in cognitive penetration processes. It is no help in responding to the challenge to simply specify that the belief must play an essential justification-related role without explaining when and in virtue of what it does so.[55]

---

[53] Counterfactual tests are not perfectly reliable tests of causal contribution. The test is used here to illustrate how a causal test for essentialness fails, not as a test for causal contribution in general.

[54] Note that the case can still be an instance of cognitive penetration of Jill's experience by her pre-experience belief that Jack is angry, even though she would have formed reaffirmed that claim without the penetrated experience. The pre-experience belief causally contributes to her having an experience with the precise phenomenology it has. The claim that Jill's pre-experience belief causally contributes to her having that precise experience passes the counter-factual test: had she not had the belief, she would not have had an experience with that precise phenomenology. Note also that we can specify that in the counter-factual non-penetration case, Jill would have believed that Jack is angry on the basis of a neurotic cause that would not have actively contributed in the penetration causal chain. Thus, Jill's penetrated experience and its improper etiology can still feature in the explanation of her belief in the actual case.

[55] The focus on essentialness may be ill-conceived for another reason. Perhaps the added condition should track the degree to which an input-belief's justification-status makes a difference to the output-belief's justification-status in the relevant cases. As a result, reliabilists could appeal to justificatory relevance of input beliefs rather than their essentialness. However, moving from essentialness to relevance does not make the task of responding to the challenge

In this sub-section, I have raised a number of problems for the possibility that reliabilists can respond to the cognitive penetrability challenge by appealing to belief-dependent justification. Appealing to belief-dependence does not address cases like Wishful Grade. In addition, reliabilists must say more about how they plan to treat a range of examples involving belief-dependence, where input-beliefs play a variety of roles. The options surveyed face a number of complications.

6 How the Challenge Interacts with other Problems for Reliabilism

Cognitive penetrability raises a new challenge for Reliabilism about justification. However, the challenge also interacts with more familiar problems for the view. In this section, I explain how the challenge interacts with the generality problem and new evil demon problem for Reliabilism.

6.1 How the Generality Problem Bolsters the Challenge

As noted above, the token process that produces a belief is not assessable for reliability; only types of process are. But token beliefs fall under indefinitely many types, and these types vary in their degrees of reliability. In order for Reliabilism to yield determinate verdicts about the justification of beliefs, there must be a unique, relevant process-type whose reliability fixes the degree of justification for the belief.

---

any easier for reliabilists. Specifying a relevance condition faces the same difficulties raised for proposals concerning input-beliefs' essential roles. Although logical, probabilistic, and causal factors may be relevant to the justification-statuses of output-beliefs in cases of cognitive penetration, it is not clear what condition captures their precise role.

The generality problem arises because it is not clear whether Reliabilism has the resources to identify a unique, relevant process-type from all of the possible candidates.

In pressing the generality problem, Conee and Feldman (1998) argue that an adequate solution must specify a general way of identifying such a process-type that is principled, consistent with our epistemic intuitions, and within the spirit of Reliabilism (cf. Feldman, 1985).

My arguments above that the cognitive penetrability challenge arises for Reliabilism are distinct from the generality problem. I have argued that Reliabilism has false entailments for one or more cases of cognitive penetration no matter which process-type was considered, for a range of process-types. That is, no matter which process-type turns out to be the relevant one, I argued, reliabilists face the challenge.

However, the considerations that give rise to the generality problem bolster the cognitive penetrability challenge to Reliabilism. Suppose that, in Jill's case, reliabilists identify an unreliable process-type which yields the true verdict. Numerous reliable process-types still compete as candidates, and reliabilists need an argument for counting their favored process-type as relevant. It is unclear what the argument would be, since the process-types I have surveyed are all plausible candidates.

6.2 How the New Evil Demon Problem Bolsters the Challenge

In one of the earliest challenges to Reliabilism, a number of authors argued that the

view falsely entails that demon-deceived subjects' beliefs are not justified (e.g. Cohen, 1984). Imagine Dan, whose experiences, memories, and other states have been implanted by an evil demon determined to deceive Dan. Dan forms beliefs about his surroundings, etc. However, most of the beliefs are false. None of the process-types that produce Dan's beliefs are reliable in Dan's world. As a result, it appears that Reliabilism predicts that Dan's beliefs are not justified. But the prediction is false; a demon-deceived subject's beliefs are justified.

The cognitive penetrability challenge is logically independent from the new evil demon problem. One can consistently reject the claim that the beliefs of demon-deceived and hallucinatory subjects are justified, for example, and still accept the key verdicts about justification that underlie the cognitive penetrability challenge. In addition, a solution to the new evil demon problem does not guarantee a solution to the cognitive penetrability challenge. However, despite the logical independence, the new evil demon problem can be used to bolster the cognitive penetrability challenge to Reliabilism. Some attractive responses to the new evil demon problem force reliabilists to accept untenable responses to the cognitive penetrability challenge. For example, some reliabilists explain the intuition that a demon-deceived subject's beliefs are justified by pointing to a general process-type in common between the demon-deceived subject and non-deceived subjects in the actual world with qualitatively identical experiences to their demon-world counterparts. Goldman (1992) argues that both subjects' beliefs are justified (or more precisely, that we count them as such)

because both sets of beliefs are (immediately) based on visual experience and visual experience is (considered to be) a reliable process-type (cf. Goldman, 2008b). If *the process-type that produces beliefs on the (immediate) basis of visual experience* is the relevant type for the demon-deceived and normal subjects, it is plausibly the relevant type for the subjects in our cases of cognitive penetration, as well. However, as noted in §4, *the process-type that produces beliefs on the (immediate) basis of visual experience* is reliable. So, if that is the relevant process-type in cases of cognitive penetration of visual experience, Reliabilism falsely entails that the subjects' beliefs in our cases of cognitive penetration are justified.[56]

6.3 Mutual Reinforcement

As I outlined above, two familiar challenges to Reliabilism can be used to bolster the cognitive penetrability challenge to Reliabilism. A more general point is this: all three of these challenges to Reliabilism—regarding generality, new evil demon, and

[56] Although the new evil demon problem can be used to bolster the cognitive penetrability challenge, one could also try to use the connection to shed light on how reliabilists can respond to the cognitive penetrability challenge. One might do so, for example, by trying to leverage proposed solutions to the new evil demon problem to yield a solution to the cognitive penetrability challenge. For example, one could follow Bach (1985) and Engel (1992) in distinguishing doxastic from personal justification. One could then argue that the key verdict about doxastic justification in the cases of cognitive penetration discussed above is false, but an easily conflated verdict about personal justification is true. On this line, in the cases of cognitive penetration under discussion, the subjects beliefs are doxastically justified but the subjects are personally unjustified in forming their post-experience beliefs. Although this provides a potential line of response, there are at least the following problems facing it. First, the claim that doxastic and personal justification can come apart in the suggested way is controversial. For reasons to resist the claim, see for example Kvanvig and Menzel (1990). Second, as noted above, the cognitive penetrability challenge has previously been formulated in terms of doxastic justification, not personal justification. The proposed response would require modifying the challenge to be in terms of personal justification. It is not clear how much of an advantage the cognitive penetrability challenge could provide for reliabilists over their rivals on that formulation, since it might then appear that the two groups were simply talking about different epistemic properties. Finally, even if one grants that doxastic and personal justification can come apart in the suggested way, reliabilists would still need to explain why we should care about doxastic justification so conceived, and not merely personal justification.

cognitive penetrability—mutually reinforce one another. The cognitive penetrability challenge highlights the fact that generality may pose a special problem for reliabilists attempting to account for the key verdict. The cognitive penetrability challenge also constrains the available solutions to the new evil demon problem or raises the costs of some otherwise attractive solutions. Thus, the cognitive penetrability challenge has broader implications and raises a more potent series of challenges for reliabilists than one might initially think.

7 Conclusion

In this chapter, I have argued that the cognitive penetrability challenge arises for Reliabilism, and I have offered a series of reasons to think that reliabilists cannot easily dismiss or solve the challenge. Reliabilists appear to be in a position similar to their opponents who have been previous targets of the challenge. On one hand, the conclusion should be surprising. Reliabilists have been prominent proponents of the challenge against their rivals on the assumption that the challenge does not also arise for Reliabilism. But if the arguments in the present paper are correct, the cognitive penetrability challenge does arise for reliabilists. At a minimum, appeals to the challenge fail to give reliabilists an advantage over their rivals, and the examples may refute Reliabilism if no adequate response can be found.

On the other hand, the result should not be wholly surprising. Cases of

cognitive penetration raise a challenge for theories of (perceptual) justification by identifying specific etiologies of perceptual experiences that are epistemically improper. Examples such as Graphological ALJ and Wishful Grade provide evidence that specific (types of) etiologies of perceptual states prevent those states from providing the relevant justification. Any theory of perceptual justification that entails that those specific (types of) etiologies can lead to the relevant justification faces the cognitive penetrability challenge. Dogmatism is a target of the challenge because it emphasizes the justificatory role of the content and phenomenology of experience and gives no justificatory role to the etiology of experience, beyond the etiology's role in fixing the content and phenomenology of the state. According to Reliabilism, by contrast, the etiology of an experience (and the resulting beliefs) plays a central role in perceptual justification. But while reliabilists might be right to give the etiology of experiences an important role in their account of justification, there was never a guarantee that they had identified all, or indeed, any of the features of the etiology relevant to justification. The central claim for which I have argued is that Reliabilism has some of these false entailments and thus faces the challenge. It remains to be seen what place, if any, reliability has in an adequate response to the challenge.

The debate about cognitive penetrability has generally focused on the divide between what we might call etiological and non-etiological theories of justification, pitting, e.g., reliabilists and others who emphasize the justificatory importance of the etiology of experience against, e.g., dogmatists and others who do not. The conclusion

for which I have argued here suggests that the challenge should be framed more broadly. The challenge from cognitive penetrability should not be case as a debate pitting etiological theories of (perceptual) justification against their non-etiological rivals. Rather, the challenge applies to a wide range of theories, including those that emphasize the role of etiology in perceptual justification and those that do not. In the end, the challenge is to spell out the subtle conditions under which the etiology of an experience affects the justification it provides.

CHAPTER 3

MORAL EMOTIONS, SOCIAL PSYCHOLOGY AND IRRELEVANT FACTORS

0 Introduction

In the past decade, empirical moral psychologists have uncovered some surprising influences on moral attitudes and decision-making. For example, Wheatley and Haidt (2005) hypnotized participants to feel disgust whenever they read a trigger word (e.g. 'take') and found that the participants made more severe moral condemnations of unrelated actions described in vignettes when the trigger word was present than did control participants. Schnall, Haidt, Clore, and Jordan (2008) found that exposing participants to fart spray on a campus lawn or seating them at a dirty desk in a lab increased the severity of their moral condemnations of others' actions. And Eskine, Kacinik, and Prinz (2011) found that participants who drank a bitter beverage made more negative moral assessments of others' actions than did participants who drank a sweet or neutral beverage.

Complementing the above results on the role of extraneously induced disgust, there have been some even more intriguing results concerning the relationship between moral attitudes and cleanliness. For example, Zhong and Liljenquist (2006, Study 3) found that participants who recalled their own immoral behavior tended to choose an antiseptic wipe rather than a pencil as a free gift for participation in a study. The results suggest that feelings of moral guilt can cause a felt need to physically

cleanse. In follow-up work, Lee and Schwarz (2010) found an even more specific link between feelings of guilt and bodily cleansing. Participants who had engaged in a role-playing task in which they told a malevolent lie via email, using their *hands* to type the message, preferred hand sanitizer to mouth wash when offered a choice. Those who lied via voicemail, using their *mouth* to tell the lie, preferred mouth wash over hand sanitizer.

The above results and many others like them raise a host of questions for moral psychology. In this chapter, I focus on their significance for moral epistemology. I assess a recent trend in which moral philosophers use results like those above to argue against *moral intuitionism*, the conjunction of the claims that that we have immediate justification for moral beliefs and immediate moral knowledge. The core idea of the challenge is this: empirical psychology reveals that morally irrelevant factors—reading hypnotically primed words, smelling fart spray, drinking a bitter beverage—can have a statistically significant, covert influence on people's moral attitudes. Because the results show that moral attitudes are sensitive to irrelevant factors, they seem to provide evidence that intuitive moral judgments are unreliable and, so, not knowledge. The full challenge develops the core idea into a sophisticated, empirically-based argument against intuitive moral knowledge.

My aim in this chapter is to respond to the most promising version of this empirical challenge to moral intuitionism, which I call the *Argument from Irrelevant Factors*. I argue that the challenge fails in its current forms and that its prospects are

dim. My response is novel in the following way. While others have responded to the challenge on behalf of moral intuitionism, they have overlooked important details in the empirical data. In particular, no one in the debate has given adequate attention to the small effect sizes in the results most commonly cited. While the results are important, they do not provide evidence of the unreliability of the intuitive moral judgments and beliefs that, according to moral intuitionists, constitute immediate moral knowledge. Having argued against the empirical challenge in the first part of the paper, I then turn the tables on the challengers by arguing that the data they cite against moral intuitionism actually help provide some much-needed empirical support for the view.

In addition to containing a more subtle analysis of the empirical data, my response has a further important difference from previous responses. Extant responses are *concessive*: defenders of intuitive moral knowledge concede that the studies show that intuitive moral judgments are unreliable in the experimental circumstances reported.[57] They concede that the data reveal that in many (experimental) circumstances intuitive moral judgments are unreliable. The only remaining question, it might seem, concerns how frequently the circumstances occur in the real world.[58] In contrast to these concessive responses, the argument in the present paper shows that the results do not provide evidence for the unreliability of

[57] See, for example, Shafer-Landau (2008), Tolhurst (2008), and Liao (2010).
[58] See Sinnott-Armstrong (2011) for this summary of the state of play. Cf. Shafer-Landau (2008).

intuitive moral judgments *even in the experimental contexts* most frequently cited. As a result, the present argument is not concessive, and it suggests that the current shape of the debate is misleading. This is good news for defenders of intuitive moral knowledge, since the relevant circumstances may be widespread.

Chapter 3 is organized as follows. In §1, I cover preliminaries. In §2, I present some additional empirical findings that are supposed to spell trouble for moral intuitionism. In §3, I formulate the most promising version of the empirical challenge that relies on these findings. In §4, I offer my response to the challenge by giving a more detailed assessment of the empirical results than has been offered in the literature. In §5, I explain how moral intuitionism could benefit from additional support and sketch an empirically plausible form of moral intuitionism according to which moral emotions play a role in grounding immediate moral knowledge. In §6, I turn the tables on the challengers by arguing that the findings that were supposed to undermine moral intuitionism (and other related results) can be used to support moral intuitionism. I conclude in §7.

1 Preliminaries

In this section, I define key terms and introduce the general contours of the debate as issue. Psychologists and philosophers have shown a great deal of interest in intuition, in general, and moral intuition, in particular, over the past decade. As the term

'intuition' is commonly used in the psychology literature, a judgment results from intuition if and only if the judgment is made suddenly and effortlessly without the subject's relying on premises to which she has conscious access.[59]

There are also philosophers' senses of 'intuition' and 'intuitionism', and these are different from the psychologists' senses of the terms. For the present paper, I will define the terms so as to mark out an important view in moral epistemology: moral intuitionism. Moral intuitionism (MI), as I will understand it, is the conjunction of the following two existential claims—one about justification and one about knowledge.

MI_J There is immediate justification for moral claims.[60]

MI_K There is immediate moral knowledge.

To clarify the theses, I should say more about the terms. One has *justification* for a claim if and only if one has an epistemic reason to believe it. One has *immediate* justification for a claim if and only one has justification for the claim, and the justification does not depend on one's justification for any other claims. Finally, one has *immediate knowledge* of a claim if and only if (i) one knows the claim, (ii) one has immediate justification, *j*, for the claim, (iii) one's immediate justification, *j*, is

[59] See Haidt (2001), Bastick (1962), and Bruner (1960) for definitions of 'intuition' along these lines

[60] We could in fact divide the view into more specific components: one for propositional justification and one for doxastic justification, where one has propositional justification for p if and only if one has good reasons to believe p and one has doxastic justification for a belief that p if and only if (i) one believes that p and (ii) one believes p on the basis of the good reasons on has. The version formulated in the text denotes propositional justification.

sufficient to satisfy the justification condition for knowledge of the claim, and (iv) one knows the claim on the basis of *j*.[61]

In §6, I argue that the empirical evidence that supposedly undermines MI does not, and in fact provides some support for the view. In addition, MI has been one of the most salient targets of the argument from irrelevant factors. These facts warrant keeping MI in plain sight throughout the discussion. However, the argument from irrelevant factors threatens to undermine a much wider range of views in moral epistemology than MI. Targets of the challenge include any view that entails that we have moral knowledge in the relevant circumstances via intuition in the *psychologists'* sense. That is because most of the results cited as support for the argument from irrelevant factors concern any moral judgments made suddenly and effortlessly without the subject's relying on premises to which she has conscious access, even if those judgments are not immediately justified or known in the philosophers' sense. As a result, the challenge should be of interest to anyone who thinks we have intuitive moral knowledge in the psychologists' sense, a group that includes most anti-skeptics about moral knowledge.

2 Further Recent Findings

[61] One might wonder why immediate knowledge is not simply knowledge that p together with immediate justification for p. The additional conditions are required to ensure that the immediate justification is essential to the knowledge in the right way. For example, the additional conditions exclude cases in which one has only a small degree of immediate justification to believe p and knows p partly because of some other mediate source of justification. The conditions also exclude cases in which one's immediately justified belief is not knowledge due to Gettier-style circumstances.

In the introduction, I mentioned some intriguing results from empirical moral psychology that may spell trouble for moral intuitionism. The results already mentioned are part of a much larger trend in moral psychology, in which (arguably) morally irrelevant factors seem to have significant effects on moral judgment. In this section, I discuss a few more of the most often cited studies.

Perhaps the most widely-discussed results pertain to *framing effects*. When an attitude or action is subject to a framing effect, it is sensitive to how the scenario or problem is framed. Kahneman and Tversky (1979) were the first to study such effects. Framing effects include wording effects and ordering effects. A number of experiments have found that moral attitudes are subject to framing effects. For example Petrinovich and O'Neill (1996) gave participants a number of variations on trolley-problem scenarios.[62] They found that participants' moral assessments of the same action changed depending on how the action was described. For example, responding to a single scenario with an identical base description, participants in different groups varied in their moral approval of acts when the acts where described variously in terms of 'saving 5 of 6' or 'killing 1 of 6'. In addition, Petrinovich and O'Neill also found that approval of identically described acts in identical vignettes changed when the order in which the vignettes were presented changed. Arguably the wording of the description of an action makes no morally relevant difference to whether the action is right or wrong. And, even more clearly it would seem, the order

in which the vignettes are presented makes no difference to the truth of moral claims about what it is right to do in the scenarios described. As a result, Petrinovich and O'Neill's results provide evidence that moral attitudes can be influenced by factors irrelevant to the truth of target moral claims (i.e. morally irrelevant factors).

Framing effects like those above have been the most widely cited in discussions of the apparent role of morally irrelevant factors in shaping moral assessments. However, additional studies have uncovered the influence of many other seemingly irrelevant factors. For example, Schnall, Benton, and Harvey (2008) found that decreased feelings disgust lessen the severity of such judgments. Participants who watched a disgusting film clip and then allowed to wash their hands made less severe moral condemnations of actions by characters in vignettes unrelated to the film than did participants who watched the film but were not allowed to wash their hands before making the moral assessments.

Extraneously induced anger can also influence moral attitudes. DeSteno, Petty, Rucker, Wegener, and Braverman (2004) found that extraneously induced anger affected participants' judgments about the goodness/badness of a proposed tax policy. Participants first read news articles about anti-American protests in the Middle East, previously shown to induce anger (DeSteno et al., 2000). Then, during what was ostensibly a second study, participants who had been induced to feel anger gave more favorable ratings to a tax policy framed in terms of preventing justice violations than

[62] For classic discussions of the Trolley Problem, see Foot (1978) and Thomson (1976, 1985).

did participants not induced to feel extraneous anger.

Positive mood can also have an effect. Valdesolo and DeSteno (2006) found that inducing a positive mood can cause subjects to make more utilitarian responses in trolley cases. Subjects who watched a comedic video were more likely to judge that the fat man should be pushed off of the bridge in the classic footbridge trolley scenario than subjects who watched a neutral video.

The results described above all have something in common: they provide evidence that factors extraneous to the moral claims at issue can influence moral attitudes toward those claims. The framing effects, extraneously induced emotion cases, and other results are representative of a larger trend in moral psychology. There have been numerous recent empirical studies in which factors extraneous to the truth of target moral claim have been shown to influence subjects' assessments of the claims.

What do these results tell us about the grounds of moral judgment? A number of recent authors have taken a pessimistic attitude. Walter Sinnott-Armstrong (2006) sums up what he takes the upshot to be for moral intuitionism and related views by comparing the ordinary situations in which we make moral judgments and a Gettier example in which the subject is in a country with numerous, undetectable barn fasçades.

[We can] compare [ordinary situations in which we form intuitive moral

> judgments to] a country with lots of barn fasçades that look just like real barns when viewed from the road (Goldman, 1976). If someone looks only from the road, then he is not justified in believing that what he sees is a real barn, at least if he should know about the barn façades. The barn façades are analogous to situations that produce distorted moral beliefs. Since such distortions are so common, morality is a land of fake barns. (2006: 362-363)

The pessimistic conclusion may appear warranted. If there is a large body of research in moral psychology that shows moral judgments are susceptible to the influence of morally irrelevant factors, then it is natural to conclude (or at least strongly suspect) that intuitive moral judgment is unreliable and fails to constitute moral knowledge. In the next section, I look at how to make this natural line of thought more precise.

3 A Cognitive Scientific Challenge to Intuitive Moral Knowledge

In the previous section, I reported a number of recent empirical findings suggesting that extraneously induced emotion can influence moral attitudes, and I gave an informal sketch of the challenge. I now formulate the challenge as a valid argument against any view that entails that we have moral knowledge of the relevant claims in the experimental conditions, with a focus on MI_K.

The Argument from Irrelevant Factors

P1. Intuitive moral judgments show a frequent, pronounced sensitivity to morally irrelevant factors.

P2. If intuitive moral judgments show a frequent, pronounced sensitivity to morally irrelevant factors, they are unreliable.

C1. Intuitive moral judgments are unreliable.

P3. If intuitive moral judgments are unreliable, intuitive moral judgments do not constitute moral knowledge.

C2. Intuitive moral judgments do not constitute moral knowledge.

P4. If intuitive moral judgments do not constitute moral knowledge, moral intuitionism about moral knowledge (i.e. MI_K) is false.

C3. So, MI_K is false.

Conclusion C3 of the argument is the one that targets moral intuitionism in the philosophers' sense. However, as the intermediate conclusions make clear, the argument also threatens other views in moral epistemology. C2 is inconsistent with any view that entails that subjects have moral knowledge via intuitive moral judgment under the relevant circumstances, whether that knowledge is immediately justified or not. And C1 threatens that claim that we obtain justified moral beliefs as the result of

intuitive moral judgments on reliabilist conceptions of justification.[63] Thus, the Argument from Irrelevant Factors seems to threaten a very wide range of attractive, non-skeptical moral epistemologies.[64]

The argument above falls under a general type of argument that has recently emerged in the moral psychology and meta-ethics literature as a prominent challenge to various kinds of moral knowledge.[65] The above version of the argument avoids several pitfalls for arguments of its type, and is worth taking seriously. Most importantly, the argument promises to yield its conclusions without requiring controversial stands on normative issues. The evidence for P1 comes from the sort of cases discussed in the previous section, in which moral emotions are sensitive to factors which are supposed to be *uncontroversially* irrelevant to the truth of the target moral claims.

That the factors in question are uncontroversially morally irrelevant to the truth

---

[63] For an exposition and defense of a process reliabilist conception of justification Goldman (1979, 1986). For a version of moral intuitionism that relies centrally on a reliabilist notion of justification see Shafter-Landau (2003).

[64] The argument is plausibly not successful on some conceptions of epistemic justification. For example, it is implausible that a successful version of MI can be formulated in terms of an *internalist* notion of epistemic justification, according to which one can justifiedly believe that *p*, even if the belief was not reliably produced. Sinnott-Armstrong (2006, 2008) seems to vacillate on whether the relevant conception of justification is reliabilist or some other conception. In his (2008), he puts the challenge in terms of reliability. But in both his (2006, 2008) he argues that the subjects targets by the argument know or should know about the relevant moral psychology results. That suggests that he has an access internalist version of the challenge in mind, on which it matters that the relevant information is accessible to the subjects. However, it is implausible that most people know or should know about these results. So on what we could call the internalist formulation, the challenge does not apply to the vast majority of people. Moreover, given plausible assumptions about the nature of defeated justification, it is not successful even for people who do find out about the experimental results, as Tolhurst (2008) points out. First, even if the results do defeat their immediate justification for all of their moral beliefs, that justification can restored if the defeating evidence is itself defeated. Immediate justification which was defeated and then restored by defeating the defeater can be immediate justification. Second, even if the restored justification were not immediate, the original justification could have been. Finally, as I argue below, the studies themselves do not provide evidence sufficient to defeat the relevant justification in the first place.

of the claims at issue promises to set the above version of the argument apart from another version of the argument, due to Joshua Greene and colleagues (Greene et al. 2001; cf. Greene 2007, 2009).[66] Greene et al. argue that certain emotional moral judgments are sensitive to factors that determine whether the actions are "up close and personal" or not. However, Greene et al's characterization of what makes an act "up close and personal" is controversial and has evolved in response to criticisms. Greene sometimes characterizes the factors in a way that requires taking controversial stances concerning the truth of consequentialist and deontological theories.[67] And it is not clear that moral judgments track the "up close and personal" factors of an act when it is characterized in a way that is clearly irrelevant in the right sense; i.e., unreliable indicators of moral truths. Insofar as it utilizes evidence concerning factors that are uncontroversially irrelevant to the truth of the moral judgments at issue, the present version of the argument from irrelevant factors is potentially a significant improvement over Greene et al.'s version.

In what follows, my main complaint will be with P1, the claim that intuitive moral judgments show a frequent, pronounced sensitivity to morally irrelevant

---

[65] Laboratory-based empirical arguments are presented in Greene et al. (2001), Sinnott-Armstrong (2006, 2008), Levy (2006), and Nadelhoffer and Feltz (2008), among others. Evolutionary arguments drawing on empirical considerations include Street (2006) and Levy (2006).

[66] I should note that Greene et al. do not aim to refute moral intuitionism in general. Their focus is rather on deontological moral theories and the emotion-based processes they think underlie support for such theories. However, since I will defend an emotion-based moral intuitionisim in the final section, it is significant for my purposes that the present version of the AIF is superior to Greene's in the way mentioned, since I will not respond to Greene at length in this paper.

[67] For relevant criticisms of Greene et al., see Kamm (2007), Berker (2009) and McGuire, Langdon, Coltheart and Mackenzie (2009).

factors. I want to focus on its formulation in this section.[68] In the next section, I assess the evidence for P1.

P1 must be formulated in terms of a *frequent* and *pronounced* sensitivity to morally irrelevant factors in order to make P2 plausible. P1 is the antecedent of the conditional P2, which says: If intuitive moral judgments show a frequent, pronounced sensitivity to morally irrelevant factors, these moral judgments are unreliable. If intuitive moral judgments showed only a rare or tiny sensitivity to morally irrelevant factors, they might well still be reliable.

To see why, consider an analogy with color vision. Vision is sensitive to factors that are irrelevant to the common-sense color of objects. For example, if a red light shines on white table in some circumstances, the table will look red. In such cases, color vision is sensitive to factors irrelevant to the common sense color of the table. But such cases are rare, and color vision is reliable despite their occurrence. The effects must be sufficiently frequent to threaten reliability. Similar points apply to the qualification that the effects must be pronounced. Color vision scientists have determined through color matching tasks that there is a fairly wide degree of variety among normal sighted humans in the precise hues their visual experiences represent for any given object. Although the visual systems of most such subjects are arguably

[68] However, I should note that other premises of the AIF could be attacked, and some of the extant responses to the argument can be seen as undermining them. For example, Shafer-Landau (2008) gives a response that could undermine P4: even if the intuitive moral judgments (in the psychologists' sense of 'intuitive' that is relevant for P4) that we see in the experimental findings are not moral knowledge, there might be other moral beliefs which are immediate justified in the philosophers' sense; e.g. the beliefs in W. D. Ross's *prima facie* duties.

unreliable in their representations of fine-grained hues (such as red_{32} and $yellow_{17}$), they are plausibly reliable in their representation of coarse-grained hues (such as red and yellow). Analogously, evidence of relatively infrequent or small effects would limit the evidence against the reliability of intuitive moral judgments, and hence limit the support for P2. The upshot is that P1 needs to be formulated in terms of intuitive moral judgment having a *frequent* and *pronounced* sensitivity to morally irrelevant factors.

4 Reply to the Argument from Irrelevant Factors

I now turn to an assessment of the evidence for P1. P1 is a claim about intuitive moral judgment in general, both in and out of the experimental contexts. Others have suggested that even if intuitive moral judgment shows a frequent and pronounced sensitivity to morally irrelevant factors in the experimental contexts, there might be no such sensitivity out in the real world. In addition, it is common to concede that the studies show at least that the kind of intuitive moral judgments in question are unreliable in the experimental contexts. The current trend in the debate suggests that empirical psychology has revealed that, in many circumstances, intuitive moral judgments are unreliable, and the only remaining question is how frequently the circumstances occur in the real world.[69]

---

[69] Sinnott-Armstrong (2011) thus recommends that moral psychologists devote more time to trying to assess how frequent the morally irrelevant influences occur.

4.1 Avoiding a Concessive Response

If the shape of the debate is as described above, anti-skeptics about intuitive moral knowledge are in serious trouble. The kind of morally irrelevant factors that can influence moral attitudes appear to be numerous and widespread. Think of how often you wash your hands, smell a slightly foul odor, feel a bit frustrated, or change moods. These factors have all been shown to have some influence on moral attitudes, and they present frequent influences. If intuitive moral judgment is unreliable in the experimental contexts, I think it is likely to be unreliable in many contexts. Thus, extant responses to the AIF suggest that moral anti-skeptics about intuitive moral judgments are in serious trouble.

However, I will now argue that the extant responses are too concessive: the experimental findings do not provide evidence that intuitive moral judgments are unreliable, not even in the experimental contexts. To see why, we first need to look at how the experiments are designed.

4.2 Notes on Experimental Design

In this subsection, I give a brief overview of the design of the experiments in question, highlighting the aspects that are most important for a non-concessive response to the AIF.

Participants in the studies read vignettes in which characters perform acts. The acts range from morally neutral (e.g. setting up discussions at a school) to morally

wrong (e.g. stealing, lying, or killing). The participants then rate the acts under various experimental conditions. For example, in one extraneous emotion study, the experimental group of participants rates the act while being exposed to a disgusting smell in the air while a control group rates the act under normal conditions. Or in a framing effect study, after reading identical vignettes, one group rates the act in the vignette under the description 'saving five of six' while another rates the act under the description 'killing one of six'.

For my purposes, it is important to look closely at how the ratings work. There are two general formats. In one format, subjects are asked to rate the acts on a continuous scale. The property spectrum on the scale differs across studies. In the disgust studies, the spectrum typically ranges from 'not at all morally wrong' to 'extremely morally wrong'. Participants are thus asked to rate the act for its precise *degree of wrongness*. In the framing studies using continuum answers, things are usually a bit different; subjects rate how strongly they agree or disagree with the statement that they would (or should) perform the action under the circumstances described in the vignette.[70] Participants in both emotion and framing studies often indicate their answers by putting a slash mark on a line segment with labels at the extremes and

[70] The presence of the word 'would' here is unfortunate. For our purposes (and usually for the experimenters' as well) the key issue is normative. The issues concern participants' moral assessments of the acts, not predictions of what they would in fact due (perhaps contrary to their own best judgment of what they ought to do) Thus normative terminology (e.g. 'should') is more appropriate. However, a meta-analysis across studies that variously use 'would' and 'should' reveals highly similar statistical results. E.g. when 75% of participants agree to the 'would' formulation, one also finds that roughly 75% agree to the 'should' formulation as well. It is possible that the 75% in the two cases represent different portions of the population sampled, but that would be quite a surprising coincidence. While more research is needed on the issue, it is thus fairly safe to assume that participants interpret the two formulations in roughly equivalent ways and

various mid-points. In the second format, subjects indicate answers to 'yes' or 'no' questions.

In the discussion below, I will focus mostly on the continuum format. At the end of the section, I explain how the conclusions I draw about the results obtained in the continuum format apply to the yes/no format.

4.3 Answers and Judgments

I this subsection, I look even more closely at the answers participants gave in the relevant studies in order to determine what, if anything, we can learn from them about the reliability of intuitive moral judgment.

4.3.1 Participants' Attitudes

Before we can assess whether it is likely that participants made moral judgments in the experimental circumstances it will be helpful to distinguish some kinds of attitudes participants might be expressing with their answers. Consider the following two kinds of judgment. The *coarse-grained* judgments that concern us are about whether an act has a coarse-grained moral property or not, where such properties include wrongness, rightness, permissibility, impermissibility and the like. For example, a coarse-grained judgment could be about whether an act is morally wrong or whether one should perform the act. *Fine-grained* judgments, by contrast, are about the precise degrees to

---

that studies using either formulation can be merged for the purposes of discussions like the present. Thanks to David Pizarro (p.c.) for helpful discussion on these issues.

which things possess the relevant properties. For example, a fine-grained judgment could be about the degree to which an act is morally wrong or the degree to which one agrees that the act should be performed.[71]

4.3.2 Do the Participants Make Judgments, and, If So, Which Judgments?

It is obvious that the participants give answers, but it is not obvious that these answers reflect participants' moral *judgments*. A judgment requires a kind of endorsement that an answer does not. The results tell us where the participants were willing to give an answer, but not how much confidence they had in their answer either at the coarse-grained or fine-grained level. A participant may have been quite unsure of where to place the mark. Compare: if you are taking a calculus test and you do not know the right answer to some question on it, you could give '1' as your answer without judging that '1' is the answer. In short: answers do not entail judgments.[72]

The worry that participants in the experimental contexts are not making judgments is especially plausible with respect to fine-grained moral judgments. For example, there is little evidence that subjects endorsed the precise degree of wrongness indicated by the placement of their slash marks. Compare: if I ask you to

[71] A further complication concerns the distinction between outright judgment or belief and degrees of belief. Participants can express degrees of belief without expressing outright judgments. Outright judgments and beliefs can be assessed for reliability, but degrees of belief arguably cannot. Some experiments seem to target subjects' degrees of belief; e.g. those that ask participants to rate their agreement with a claim. Others seem to target the subjects' outright judgments, e.g. those that ask subjects to give an answer about whether an act was right or wrong. For present purposes, I suppress this additional complication. I note it here because it raises further questions about whether the results usually cited in support of the Argument from Irrelevant Factors do indeed provide support.

rate the precise, fully determinate shade of red an apple is and you match it to a swatch of red_{21}, it is unlikely that formed a judgment that it the apple is precisely red_{21}. Much more likely, you identified red_{21} as the best candidate without committing yourself to such a precise judgment. However, it is more plausible that you formed a judgment about the *coarse-grained* color of the apple; i.e. you plausibly judged that it was red.

In sum: it is clear that participants in the moral psychology studies in question gave answers, but unclear whether they formed judgments. And, regarding judgments, it is much more plausible that they formed coarse-grained judgments than fine-grained ones.

4.3.3 Answer Scales, Effect Sizes and Unreliability: the Details

Suppose that participants in the studies do make the relevant moral judgments. In this section, I argue that even on this assumption, the studies fail to provide evidence that coarse-grained intuitive moral judgments are unreliable.

4.3.3.1 Framing Effects in Detail

Let's first look at the details of the studies on framing effects. The main concern will be to assess whether they provide evidence of shifts in coarse-grained moral judgment due to the role of framing effects. I cannot discuss all the studies in detail. For ease of exposition, I will focus on results that Sinnott-Armstrong (2008) makes central to his

[72] Bengson (forthcoming) argues at length for the plausibility of a related point in response to empirical attacks on the reliability of intuitions. He takes intuitions to be 'seemings' that are distinct from judgments, but many of his points apply equally to what I have called intuitive judgments.

version of the argument from irrelevant factors. These results are representative of larger trends and other studies contain similar details. So the points made here generalize to other attempts to use framing effects to impugn the reliability of intuitive moral judgment.

A closer look at Petrinovich and O'Neill's (1996) study reveals that found no coarse-grained shifts due to framing effects for wording and ordering. The ordering aspect of the study included three pairs of forms. For each form, one group was given vignettes in one order and the other group was given them in reverse order. Participants' ratings shifted in how strongly they approved of some action, but crucially they *approved of the action* under all of the circumstances tested. For example, when the order of trolley problem vignettes was varied, participants approved of the target action with a rating of +1.0 and +2.6, on a scale ranging from -5.0 to +5.0. The ratings were confined to the positive side of the approval scale; hence the results do not provide evidence that the participants changed their coarse-grained moral judgments about the case. Similarly small effect sizes confined to the positive side of the scale were found for one of the other forms, while no statistically significant shifts at all were obtained for the third.

For some experiments reported by Petrinovich and O'Neill, there were shifts from one side of the spectrum to another for framing effects. However, the effect sizes were too small to provide evidence for unreliability at the coarse-grained level of judgment. The studies asked participants to rate their degree of agreement or

disagreement on whether they would (read: should) perform some action. The most variation spanning agreement to disagreement Petrinovich and O'Neill found showed a very slight shift from agreement to disagreement concerning the target moral act: in one study of wording effects ('killing' vs. 'saving') there was a shift from +0.65 to -0.78. The ratings were made on a scale of +5.0 to -5.0, and participants were informed that +1.0 indicated 'slight agreement' and -1.0 indicated 'slight disagreement'. The results do not even show a shift from slight agreement to slight disagreement.

As noted above, it is not entirely clear how to translate the scaled numbers to conclusions about moral judgments. If anything, this is a problem for the proponent of the AIF, since they are the ones under an obligation to provide evidence of the unreliability of such judgments. However, if we work with a plausible set of assumptions, the evidence fails to support P1. It is plausible that the space on the scale between 'slight agreement' and 'slight disagreement' does not represent participants' outright coarse-grained judgments. That is, if someone rates their agreement as 'slight' on whether one should perform an act, they plausibly have not judged that one should perform the act. In the jargon, their degree of credence falls short of outright judgment or belief. Thus, while the possible differences in fine-grained moral judgment due to framing effects do sometimes span the divide between agreement and disagreement, the effect sizes are not large enough to provide evidence of differences in outright coarse-grained moral judgment.

As noted, the above studies (and the data mentioned) are central to

Armstrong's version of the AIF. As support for P1, they are extremely weak. They plausibly do provide evidence that fine-grained moral judgments show a pronounced sensitivity to morally irrelevant framing effects. But this is not surprising. No one should have expected moral judgments to be reliable at the fine-grained level. And no one should have claimed that we have intuitive moral knowledge of fine-grained claims concerning the precise degree of an act's moral wrongness.

The claims that anti-skeptics about intuitive moral knowledge care about are coarse-grained. And the studies provide very little, if any, evidence that coarse-grained moral judgments are unreliable due to a pronounced sensitivity to framing effects. In fact, the studies seem to pain a rather different picture: because the effect sizes due to framing effects are small, the studies arguably provide evidence against P1, since they seem to reveal that coarse-grained moral judgments are fairly insensitive to framing effects.

4.3.3.2 Extraneous Emotion in Detail

Having assessed the details of some of the representative studies on framing effects, I now turn to an assessment of the results from studies on the effects of extraneously induced emotion. In these studies, we find similarly small effect sizes and the same conclusion is warranted: the empirical results do not impugn the reliability of coarse-grained intuitive moral judgments. For ease of exposition, I will again highlight a few representative results from Wheatley and Haidt (2005) and Schnall, Haidt, Clore, and

Jordan (2008).[73]

In the set-up for Wheatley and Haidt's (2005) study, participants rated the wrongness of an act by making a slash mark on a 14 cm line segment. One end of the line segment was labeled "*not at all morally wrong*" and the other end was labeled "*extremely morally wrong*". The experimenters do not specify where on the scale to locate the divide between "not wrong" and "wrong". Slashmarks on the line are converted to a scaled score from 0 to 100.

The key question for our purposes concerns whether uncontroversially irrelevant emotional factors cause subjects to make false coarse-grained moral judgments that they would not otherwise make. To assess answers to that question, at a minimum, we need to have some idea of how to map the fine-grained responses participants in the studies give to coarse-grained verdicts. That is, we need to have some idea of roughly where the dividing lines are between judgments that an act is wrong and the lack of such judgments. The lack of such judgments divides separately into judgments that the act is not wrong and suspension of judgment on the issue.

The lack of clarity on where to divide the answers between coarse-grained judgments poses an additional problem for attempts to read off results concerning subjects' coarse-grained judgments: it is not clear how to map answers on the scale to coarse-grained judgments about whether the act was wrong. In particular, it is not

[73]Sinnott-Armstrong (2006) uses Wheatley and Haidt (2005) as part of his version of the argument from irrelevant factors. Both studies are representative of typical results in the literature.

clear whether there are any cases where extraneously induced disgust caused a subject to falsely judge that an act was morally wrong when, without the disgust condition, they would not have done so. Without more clarification on the issue, the studies do not provide clear evidence of unreliability at the level of coarse-grained judgment.[74]

One way to try to locate the relevant lines is to locate the minimum threshold for judgments of outright wrongness. If we could do that, we could check whether subjects ever judge that an act is wrong only after being exposed to an extraneous influence, using control subjects as the contrast case. Since the experimenters provide no indication of where to draw the lines, though, we have to use some other methods to locate them. The most promising option is to look at the ratings that participants in the non-disgust control groups gave for a range of acts and match the ratings to one's own intuitions about the moral wrongness of the acts. The method requires using one's own intuitions about the moral wrongness of the actions, so it is not ideal. The fact that this is the most promising option available presents yet a further challenge to the argument from irrelevant factors, and suggests that emotional moral intuitionists have little to fear from the data.

Most of the vignettes used in the studies are of acts that are likely to be counted as at least somewhat morally wrong. To get a rough idea of the various ratings, consider the following data from two representative studies. In one experiment,

[74] To the extent that it is important to establish whether the coarse-grained judgments are reliable in the circumstances, the points here suggest that there could be a benefit in altering the experimental design concerning the form in which answers are reported.

Wheatley and Haidt (2005) obtained the following scaled scores (see Table 1).

Table 1	Data from Wheatley and Haidt (2005) Experiment 2					
Act	Littering	Bribery	Ambulance-chasing lawyer	Shoplifting	Library theft	Student Council
Mean rating/100 in non-disgust condition	64.71	78.88	70.39	73.06	69.53	2.7
Mean rating in disgust condition	67.64	83.86	75.37	74.34	66.14	14.0

Note that the vignettes with acts likely to be perceived as at least somewhat morally wrong all have scaled scores in the non-disgust condition about 65 and above. Student Council was a morally neutral vignette and received a scaled score of 2.7. The vignette read as follows:

> **Student Council** Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion. (Wheatley & Haidt, 2005)

Schnall, Haidt, Clore and Jordan (2008) included vignettes of acts in morally gray areas. For some of the relevant results, see Table 2.[75]

[75] To allow for more direct comparison with Wheatley and Haidt (2005), scaled scores presented here are converted from the 1 – 7 point Likert scale used in Schnall et al. (2008), where "high scores indictate[d] permissibility; low scores

Table 2	Data from Schnall, Haidt, Clore and Jordan (2008), Experiment 1		
Act	Sex between first cousins	Driving	Film
Mean rating/100 in non-disgust condition	61.85	21.71	49.42
Mean rating in weak disgust condition	72.86	28.86	56.71
Mean rating in strong disgust condition	65.71	29.29	62.43

Driving arguably includes a morally neutral act. It reads as follows:

> **Driving** James is going to work and considers whether to walk the 1½ miles or to drive in. He is feeling lazy and decides to drive in. How moral or immoral do you, personally, find James's decision to be? (Schnall, Haidt, Clore, & Jordan, 2008)

Film arguably includes an act in a morally gray area. It reads as follows:

> **Film** Controversy has erupted over a documentary film about Mexican immigrants. The film has received excellent reviews, but several of the people interviewed in it have objected that their rights were violated. The filmmaker deliberately had his camera crew stand back 15 feet in a crowd so that some interviewees did not realize they were being filmed. Because the camera was

indicate[d] moral condemnation. Note that the direction here is reversed: In Table 2, higher scores indicate more moral condemnation by participants.

not hidden, the procedure was legal. What do you think about the studio's decision to release this film, despite the aforementioned allegations? (Schnall et al., 2008)

A rough estimate of where to mark the threshold for answers that indicate outright moral wrongness seems to be as follows: ratings about 60 or 65 and above can be tentatively interpreted as indicating outright wrongness in varying degrees. Ratings somewhat below 60 or 65 indicate acts in a morally gray area. And ratings far below 60 or 65 indicate acts that are not morally wrong.

Using these very rough estimates, further analysis of the data reveals that none of the results show clear evidence of differences in coarse-grained judgment due to extraneously induced emotion. In fact, most of the experiments are not designed to test for such a shift. As noted above, most of the experimenters' vignettes involve (what the participants likely take to be) morally wrong acts. So the differences between subjects who are influenced by extraneous factors and control subjects in the typical experimental conditions show shifts in the *fine-grained* degree to which they say acts are wrong and not shifts at the coarse-grained level of judgment (e.g. from judgments that the act is not wrong to judgments that it is wrong). The above result is the result of experimental design: experiments involving extraneous emotional elicitations typically do not typically test participants' reactions to morally neutral cases. The experiments are not designed to test for evidence that such factors affect

the reliability of moral judgments at the coarse-grained level.

Comparing answers about morally neutral acts or acts in a morally gray area also does not suggest shifts in judgment at the coarse-grained level. For example, in Wheatley and Haidt's (2005) study, participants' ratings of Dan's clearly morally neutral acts in Student Council did rise from 2.7 in the non-disgust condition to 14.0 in the disgust condition. A scaled rating of 14.0 plausibly falls far short of indicating that the act was outright wrong. So it is not plausible that subjects' coarse-grained moral judgments about Dan shifted as a result of the extraneously induced disgust, even if we assume they made coarse-grained judgments. Similarly Schnall, Haidt, Clore, and Jordan (2008) found participants' ratings of James's lazy act in Driving rose from 21.71 to 28.86 (mild disgust condition) and 29.27 (strong disgust condition). These ratings are, again, likely to be short of an outright judgment that James' act was morally wrong. Finally, participants' ratings of the filmmakers' acts in Film shifted from 49.42 in the non-disgust condition to 56.71 (mild disgust condition) and 62.43 (strong disgust condition). These data suggest that participants' attitudes toward the acts in Film *might* shift from suspension of judgment to an outright judgment. For example, the strong disgust rating was about even with mean non-disgust ratings of sex between first cousins. However, the evidence is not strong, for even sex between first cousins is plausibly in a morally gray area for many people (marriage between first cousins is legal in 21 US states and Washington DC, and sex between first cousins is legal in another 10 states). Likewise, a rating of 62 is plausibly roughly within the

morally gray area for many participants.

4.4 Yes/No Formats

In the preceding discussion, I have focused on experiments involving answers on a continuum. I mentioned at the beginning of the section that some studies have a format in which participants' are asked to give 'yes' or 'no' answers about which actions to perform. The experimental design in those studies is otherwise identical to the continuum-answer studies we have investigated. Absent further evidence to the contrary, we can reasonably conclude that participants' moral attitudes in the yes/no studies also reflect only small shifts at the fine-grained level due to irrelevant factors and no regular shifts at the coarse grained-level of judgment. As a result, these studies do not threaten moral intuitionism any more than those assessed in detail above.

4.5 Diversity

The effect sizes reported above are averages of the entire samples in the studies cited. Small averages can mask pronounced effects that occur only for a small portion of a sample. It is consistent with the data discussed so far that some people's coarse-grained moral judgments are highly susceptible to the morally irrelevant factors cited in the studies. The point is important in our assessment of the AIF. If some people's coarse-grained moral judgments are highly susceptible to irrelevant factors, then their moral judgments might not be reliable and hence might not be known. Moreover,

evidence from the studies might provide evidence that defeats most people's justification for the moral judgments in question, unless and until they can determine whether they are among the group that is susceptible. And, in fact, we do find some variation in susceptibility to the irrelevant factors cited. For example, Schnall, Haidt, Clore, and Jordan (2008) found that individuals who score high in private body consciousness (PBC) are significantly more susceptible the influence of extraneously induced disgust than are low PBC individuals (no significant differences for other variables). However, these findings do not threaten my non-concessive response to the AIF. Even for the cohort that is most susceptible to extraneously induced disgust (high PBC subjects) the effect sizes were too small to suggest that shifts in coarse-grained moral judgments due to irrelevant factors are sufficiently common: the effects sizes were a only slightly over 1.0 and 0.5 on a 6-point scale for the two experiments assessing high vs. low PBC. Figure 1 is from Schnall et al. (2008) Experiment 2. Figure 2 is from Schnall et al. (2008) Experiment 3.

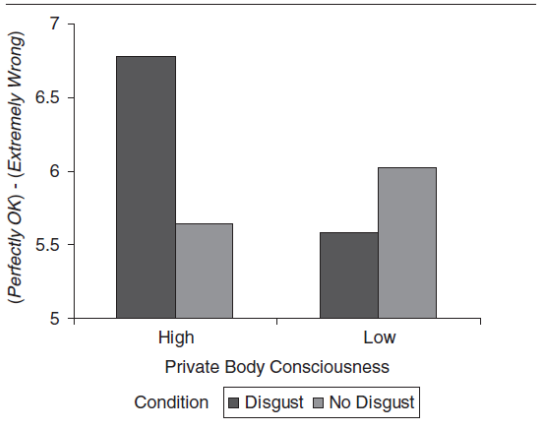


Figure 1 Judgments of wrongness of moral actions as a function of Private Body Consciousness and condition (Experiment 2).

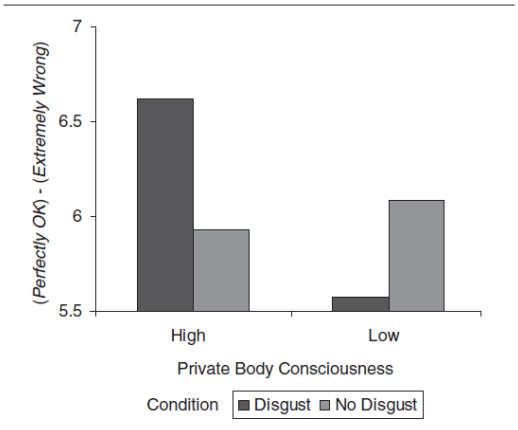


Figure 2 Judgments of wrongness of moral actions as a function of Private Body Consciousness and condition (Experiment 3).

4.6 Frequency

Above I argued that the studies commonly cited in support of P1 of the Argument from Irrelevant Factors do not support P1. I focused specifically on whether the effects in question were shown to be sufficiently pronounced to raise doubts about the reliability of coarse-grained moral judgment. An additional point worth emphasizing now concerns frequency. Recall that P1 must be formulated both in terms of the size of the influence of irrelevant factors on moral judgment and its frequency. I have not surveyed all of the experimental results here. Even though I found no evidence of effects causing differences at the coarse-grained level, I do not rule out that they exist. The response I am offering is compatible with the claim that the irrelevant factors sometimes have an influence on coarse-grained judgment. The key point for present purposes is that the experimental results used in to support the AIF do not suggest that effects at the coarse-grained level are common. Thus, even if there are sometimes coarse-grained effects, there is nothing in the data to suggest that they are common enough to impugn the reliability of intuitive moral judgment.

4.7 Summing up the Non-concessive Response to the AIF

In the above section I assessed purported evidence for the unreliability of intuitive moral judgment and found no such evidence. The evidence was supposed to come from empirical studies in which participants answers concerning the moral properties of acts described in vignettes were different as the result of morally irrelevant factors.

The differences found in these studies are important for what they tell us about the role of such factors—especially emotion—in shaping moral judgment. But the evidence poses little or no threat to plausible views that entail intuitive moral knowledge. The differences found in the studies were small, and they provide evidence against the existence of intuitive moral knowledge of fine-grained moral claims about, e.g., the precise degree to which an act is wrong. But we did not need evidence from empirical psychology to tell us that we lack such precise moral knowledge. Concerning the much more plausible claim that we have intuitive moral knowledge of coarse-grained claims concerning, e.g. whether an act is wrong, the studies do not pose a threat. In sum, a careful analysis of the details of the experimental findings reveals that they do not provide motivation for concessive responses to the AIF: in the face of such evidence one need not concede that the relevant (coarse-grained) intuitive moral judgments are unreliable, not even in the experimental contexts. This is good news for moral anti-skeptics, since the contexts like those in the experiments may be fairly common.

5 Traditional Support for Moral Intuitionism

In the remainder of the paper, I aim to turn the tables on proponents of the Argument from Irrelevant Factors. I do so by arguing that findings from empirical moral psychology, including some of those mentioned above, provide support for

moral intuitionism. In this section, I explain some shortcomings of one of the most common ways to argue for moral intuitionism; I thereby show how additional empirical support for MI is important. In the next section (§6), I present an empirically plausible version of moral intuitionism. And in the section after that (§7), I show how the empirical results discussed above can help overcome some of these shortcomings by providing additional support for the view.

5.1 The Standard Argument for MI

One traditional way to support MI is via what has been called 'the standard argument' for moral intuitionism.[76] The standard argument combines versions of three widely accepted epistemological views as premises: foundationalism, the autonomy of ethics, and moral anti-skepticism. The views are supposed to jointly entail MI. Here is one formulation of the standard argument, due to Pekka Vayrynen.

The Standard Argument

1. If we have any ethical [i.e. moral] knowledge, then such knowledge is either (a) non-inferential [i.e. immediate] or (b) based on reasonable [i.e. knowledge-yielding] inference from partly ethical premises, or (c) based on reasonable inference from entirely non-ethical premises.

[76] See Sturgeon (2002) and Vayrynen (2008) for this terminology.

2. *The autonomy of ethics*: There is no reasonable inference (deductive or non-deductive) to any ethical conclusion from entirely non-ethical premises.
3. Therefore, if we have any ethical knowledge, then such knowledge is either (a) non-inferential or (b) based on reasonable inference from partly ethical premises.
4. *Foundationalism*: If we have any knowledge (a fortiori, any ethical knowledge) that is inferential, then all such knowledge is ultimately based on reasonable inference from some knowledge that is non-inferential.
5. Therefore, if we have any ethical knowledge, then some of it is non-inferential.
6. *Ethical [anti]-skepticism:* We have some ethical knowledge.
7. Therefore, some of our ethical knowledge is non-inferential [i.e. immediate]. (Vayrynen, 2008: 491)

As formulated, the argument aims to establish moral intuitionism about knowledge. A similar argument can be formulated for moral intuitionism about justification. Although the standard argument is an influential route to moral intuitionism, it is probably not the best way to argue for moral intuitionism. Here I mention three shortcomings of the argument.

First, while the argument's premises are plausible, they are not uncontroversial. Many moral epistemologists have denied foundationalism in favor of coherentism. Moral theorists have also denied the autonomy of ethics, albeit less frequently than

they have denied foundationalism. If moral intuitionism is motivated using the standard argument, then support for the view is only as strong as support for the conjunction of foundationalism, the autonomy of ethics, and the other premises required.

Second, not only are the views represented in the standard argument open to challenge, the formulations of them required for the standard argument are even more vulnerable. There are *versions* of foundationalism, the autonomy of ethics, and moral anti-skepticism that are compatible with the falsity of moral intuitionism, and these vare among the most plausible versions of the views.[77]

Third, the standard argument fails to specify a plausible candidate for source of immediate justification for moral claims or for immediate moral knowledge. As a result, the standard argument is in many ways only a promissory note. Unless one can

---

[77] In fact, Vayrynen's formulation of the argument helps to illustrate the point. Although Vayrynen claims that the argument is valid as formulated, it is not. To see why not, consider the following schematic example, where 'E' stands for *e*thical, 'N' for *n*on-ethical, '$_c$' for *c*onclusion, and '$_b$' for *b*ase.

> S believes some ethical claim E_c, and has justification for it on the basis of immediate justification for exactly two further beliefs she holds: a non-ethical belief, N_b, and an ethical belief, E_b, Suppose that S knows N_b but does not know E_b (say, she has propositional justification for p but does not believe it).

Vayrynen's formulation of the standard argument is consistent with the claim that S *knows* E_c under the conditions described, in which case 6 would be true, because S would have some moral knowledge. However, the specifications of the case are consistent with the claim that none of S's moral knowledge is non-inferential. First, nothing in the set-up precludes (1) from being true, so let's suppose that it is. In addition, S's inference to E_c includes an ethical claim, E_b, as a premise, so the autonomy of ethics (2) is not violated. Similarly, S's inferential moral knowledge that E_c is based on a knowledge-yielding inference from a base that includes E_b, so 3 is also not violated. Finally, the justification S has for her belief that E_c is ultimately based on some immediately justified knowledge (namely, N_b), so foundationalism (4) is also not violated. The above description denotes a possible scenario in which all of the premises in Vayrynen's argument are true, but the conclusion is false. So the argument is not valid. The point is important. One possibility that moral intuitionist consistently overlook is that one has immediate justification for moral claims about which one does not form beliefs. On this possibility, the justification one gets for such claims may play a crucial role in supporting one's justification for other claims which one does believe. Knowledge that *p* entails belief that *p*. So, if one never forms a belief in the immediately justified moral claim, one will lack immediate moral knowledge. For all that, however, one might still know many moral claims, and the relevant justification may trace back to immediately justified moral claims. This complicates things for MI. It may turn out that MI_J is true while MI_K false, because subjects do not in fact ever form beliefs in the moral claims for which they have immediate justification. However, for simplicity I suppress the

offer a plausible candidate for the source of immediate justification, one's moral epistemology remains seriously incomplete and embracing moral skepticism remains a viable option.[78]

5.2 Empirically Plausible Moral Intuitionism

In this subsection, I want to fill in some of the details concerning the role emotions might play in an empirically plausible moral intuitionism. There are a number of similarities between emotions and perceptual states, and these considerations can be used in support of a form of moral intuitionism that gives emotion a significant role.[79] Consider a case of disgust.

> **Disgusting Egg** Ben takes his first trip to Southeast Asia. He has never heard of *balut*. When he is offered the egg, he is instantly disgusted by the sight of the boiled duck embryo next to the yolk. He quickly comes to believe that the egg is gross.

There are several points of similarity between Ben's state of disgust and a visual state. First, both states plausibly have what we can call *presentational phenomenology*. Like a

---

proposal in the main text.

[78] Indeed, some of the central debates in recent moral epistemology revolve around challenges to moral anti-skeptics to provide detailed, plausible accounts of which moral claims serve the crucial function of bridging the gap between fact and value, non-moral and moral, etc. To the extent that the standard argument leaves the relevant epistemology under-described it fails to answer the fundamental moral epistemic challenge.

visual state, Ben's state of disgust presents the world to Ben as being a certain way. The exact content of the state is not obvious (just as it is not obvious what the exact contents of visual states are). Plausibly the presentation in both cases is demonstrative; each state presents the world as being like *that*. In addition, both emotional and visual states seem to play a role in providing justification for claims about the way the world is. Ben does not simply come to believe that the egg is bad to ingest, he *justifiedly* believes that claim on the basis of his disgust experience. In what follows, I will suggest that a similar story can be told about instances of other emotions such as anger, and including moral emotions such as moral disgust, moral indignation and others.

In addition to having the right sort of phenomenology and justificatory role, a number of emotions are sensitive to and co-vary with moral properties. These are included in the set of so-called 'moral emotions'. The moral emotions include 'self-critical' emotions such as shame, embarrassment, guilt and 'other-critical' emotions such as contempt, anger, and disgust. In an important paper, Rozin, Lowery, Imada, and Haidt (1999) provided evidence that the "CAD Triad" of other-critical moral emotions—*c*ontempt, *a*nger, *d*isgust—tend to correlate with violations in each of three types of moral systems: those pertaining to *c*ommunity, *a*utonomy, and *d*ivinity (Schweder, Much, Mahapatra, & Park, 1997). Contempt tends to be elicited by

[79] I addressed these similarities at greater length in Chapter 1 of my dissertation. In particular, I argue that emotional and perceptual states can have similar types of phenomenology and etiology.

violations of communal and hierarchical norms. Anger tends to be elicited by violations of justice and individual rights, and by harm. And disgust tends to be elicited by violations of moral and social purity.[80]

These data provide some evidence that emotions can represent evaluative, moral properties. It is plausible that if some emotional state represents a moral property *m,* and does so with the right sort of affective phenomenology and etiology, that state can provide immediate justification for some moral claim. Moral intuitionists can appeal to these additional results to argue that moral emotions provide promising candidates as the grounds for immediately justified moral judgment and—on the assumption that such judgments are sufficiently reliable—immediate moral knowledge.

We can illustrate the plausibility of a version of moral intuitionism that gives a role to the other-critical moral emotions with a well-known example due to Gilbert Harman.

> **Harman's Cat** If you round the corner and see a group of young hoodlums pour gasoline on a cat and ignite it, you do not need to *conclude* that what they are doing is wrong; you do not need to figure anything out; you can *see* that it is wrong. (1977: 4, emphasis original)

[80] Subsequent research has suggested even more nuanced ways of dividing up the targets of the other-critical moral emotions. See, for example, Hutcherson and Gross (2011).

The example is an attractive candidate for moral intuitionists to appeal to. It is plausible that Harman's subject has justification to believe and knows that the hoodlums' act is wrong. Some moral epistemologists take Harman's talk of 'seeing' literally and defend the claim that one can literally see moral properties or facts. That is, they defend the claim that visual experience can represent moral contents, such as the property *moral wrongness* and the claim *that the act is wrong*.[81] On such a view, one can then argue that by literally seeing that the act is wrong, the subject enjoys immediate justification for the claim that the act is wrong. However, views on which we have literal moral vision are unpopular and empirically implausible. As John McDowell (1998) argues, talk of moral vision is probably best kept metaphorical.

Emotion-based moral intuitionists have a more plausible story to tell about the case. They can argue that the subject's emotional state of moral anger or indignation (re)presents the hoodlums' acts as being morally offensive. They can then argue that the moral emotional state provides the subject with immediate justification to believe that the act is morally offensive in virtue of its having that representational content and the related presentational phenomenology.

Over the past decade, psychologists have found increasing evidence of a pervasive role of emotion in moral judgment.[82] While some have arguably exaggerated

[81] See, for example, Watkins & Jolley (2002), Vayrynen (2008), and McBrayer (forthcoming); cf. McGrath (2004).
[82] For two seminal instances of this trend, see Greene et al. (2001) and Haidt (2001); cf. Allman and Woodward (2008).

the role of emotion and underrepresented the role of reason,[83] it is plausible that emotion plays a frequent and substantial role in shaping and grounding moral judgment. As a result, examples of emotion-based moral judgments provide a promising set of candidates for immediate moral knowledge.

6 Turning the Tables: Empirical Results Supporting Moral Intuitionism

The findings presented in §§1-3 are typically thought to present a challenge to moral intuitionism. In §4, I argued that the challenge fails. In §5, I presented an empirically plausible moral intuitionism. In this section, I argue that further empirical findings, including those that were supposed to undermine moral intuitionism, may in fact provide further *support* for moral intuitionists by helping intuitionists answer a formidable type of challenge.

6.1 The Dependence Challenge

The type of challenge I have in mind relies on the following key claim.

> **Dependence** We typically come to form moral beliefs by inferring them from at least some non-moral claims.[84]

[83] See Haidt (2001) for an instance of such exaggeration and Pizarro and Bloom (2003) for a critique of Haidt (2001) that emphasizes compatible roles for emotion and reason in shaping moral judgment.

[84] The name 'Dependence' is mine, but the view is widely held. For a particularly vivid, if perhaps confused, discussion which looks like it takes something like Dependence for granted, see J.J. Thomson's discussion of the problem of bridging the relevant gap from non-moral to moral in Harman and Thomson (1996). See also Shaver (1985) and Weiner (1995) cited in Cushman et al. (forthcoming) for endorsements of a normative and descriptive version of the view.

Dependence is plausible for moral beliefs in just the sort of way that a similar thesis is plausible for beliefs about the future (where our beliefs about the present and past serve as the inference base) and beliefs about unobservable entities (where beliefs about observable entities provide the inference base).[85]

For examples of the Dependence thesis, consider the following 'traditional'[86] claim that a consequentialist using their view as an inferential tool[87] would typically reason as follows: if someone S caused others harm, then S probably did something morally wrong. Or, again on the traditional account, a deontologist would typically reason as follows: if S intended to treat another as a mere means to an end, then S did something morally wrong. In short, a common view about cases of moral belief formation is that they typically must be inferred from at least some non-moral beliefs.

Dependence is a claim about psychological processes of inference; endorsing it is not sufficient to support an argument against Moral Intuitionism, which is a claim about immediate epistemic justification (support) for moral claims and immediate moral knowledge. We could (psychologically) infer moral claims from various other claims, yet enjoy immediate justification for (and knowledge of) the target moral claims nonetheless. As a result, a challenge to MI cannot rely solely on Dependence.

---

[85] See Sturgeon (2002) for more on comparison between knowledge of moral claims and claims about the future and un-observables.

[86] See Cushman et al. for the claim that this view is the 'traditional' view. The traditional view is, I think not only descriptive (as is Dependence) but also *normative*: we not only do make these inferences typically, but that's what we (epistemically) *should* do.

[87] Note that Consequentialism as a view about the metaphysics of value does not entail any view about which decision tool is morally required or pragmatically optimal. But it certain *can*, and probably often is used that way, at least in a crude form.

Instead, we should point to a related epistemic thesis:

> **Dependence*** We typically form inferentially (mediately) justified moral beliefs as the result of inferring them from at least some non-moral claims.

Although Dependence* is logically stronger than Dependence, it has a similar degree of plausibility. The challenge from Dependence* against MI proceeds as follows. Consider a proposed candidate for immediate moral knowledge. Given the Dependence thesis, it is likely that the claim is in fact inferred from some non-moral claims. And, thus, given Dependence*, it is justified (or known) inferentially, if it is justified (or known) at all.

The force of the challenge is that the plausibility of the Dependence and Dependence* theses puts a burden on moral intuitionists to provide clear evidence that their proposed candidates for immediately justified moral belief and knowledge really are immediate, since there is a general presumption that they are not. That is, while MI is compatible with both theses, Dependence* adds to the difficulty of establishing that some moral judgment is immediately justified or known.

We can see an application of the Dependence challenge by again considering Harman's Cat example. Given Dependence, it is plausible that Harman's subject quickly and unconsciously infers the moral wrongness of the act. Rather than a case of literal 'seeing' the case may be more similar to a case in which one quickly infers that

one's neighbor is not home by seeing that their mailbox is overflowing. Moral intuitionists—including those who appeal to emotions as the immediate basis of the alleged justification—are thereby under pressure to support the claim that their proposed candidates for immediate moral knowledge cannot be easily handled as cases of unconscious inference and, as such, inferential justification.

6.2 Responding to the Dependence Challenge

In this section, I develop two lines of empirical response to the Dependence challenge. One line of support comes from the data that were supposed to undermine MI. As a result, the argument presented here turns the tables on some who raise an empirical challenge to MI.

6.2.1 On Dependence: The Knobe Effect

I begin with work by Joshua Knobe on what is sometimes called the 'Knobe effect' or 'side-effect effect'. Knobe and his collaborators have found in numerous studies that in some cases when subjects make a moral judgment or form a moral attitude of approval or disapproval, their moral attitude can have an effect on their judgments about causality, intentionality, and a host of other properties that are usually thought to be non-moral.

In the most famous example, Knobe gave subjects one of either of a pair of vignettes. One vignette had the word 'harm'; the other replaced it with the word 'help'. Here's the relevant vignette:

> The vice president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm/help the environment.'
>
> The chairman of the board answered, 'I don't care at all about harming/helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed/helped.[88]

One group of study participants was asked whether the chairman intentionally harmed the environment; the other was asked whether he intentionally harmed the environment. Of the participants in the 'harm' vignette group, most answered that the chairman did intentionally harm the environment. But of those in the 'help' group, most answers that the chairman did not intentionally help the environment.

There is evidence that the Knobe Effect occurs for large number of terms (and/or concepts) other than the term 'intentionally' and its cognates. Follow-up work suggests that subjects' moral attitudes affect their attitudes about claims involving a whole host of terms (or the concepts they denote, or both) including: 'intentional action', 'causing', 'doing' vs. 'allowing', 'desiring', 'deciding', 'advocating',

'knowledge', 'happiness', 'valuing', and others.[89]

The above results provide the material for part of my response to the Dependence challenge. If participants' moral attitudes about the chairman's blameworthiness affect their attitudes toward claims about the chairman's intentions, causal role, etc., then participants' attitudes about the chairman's blame are likely to be causally prior to their attitudes toward the other claims. But then it is difficult to see how the participants' moral judgments concerning the other claims could provide the epistemic basis for their judgments concerning the chairman's blameworthiness. Claims about intentions, causes, etc. are the ones typically cited in the justification base for the relevant moral claims, according to the 'traditional' account that helps underwrite the Dependence challenge. So, Knobe's data help undermine the Dependence challenge in the relevant cases. That is because Knobe's data suggest that the candidates for the non-moral claims on which our justification for moral claims is typically supposed to rest in the relevant cases, according to Dependence*, are not the inferential or epistemic basis for the relevant moral claims after all.

6.2.2 On Dependence: Extraneously Induced Emotions

In the previous sub-section I argued that evidence for the Knobe effect helps undermine a prominent challenge to moral intuitionism. In response, one might have the following worry: perhaps there is no inference of the traditional sort (e.g. from

[88] See Knobe (2003) for the original case.
[89] For a review of this body of literature, see Knobe (2010).

claims about intentions or causes to moral claims); there might still be an inference from *other* non-moral claims. Similarly, although there may not be epistemic dependence on non-moral claims about intentions and causes, there might still be epistemic dependence on claims about other non-moral properties. In short, even if Knobe's data undermine one traditional way of filling in the details for Dependence and Dependence*, perhaps that reveals only that the traditional account had the specific details of the epistemic dependence of moral claims wrong.

In order to further respond to the challenge from Dependence*, a second line of response can be motivated by appeal to the data concerning extraneously induced emotions. In this section, I argue that these data undermine the claim that subjects typically incorporate evidence concerning the non-emotional, non-moral properties of the actions they assess, as the Dependence theses imply. As a result, the data provide additional grounds for the alternative account on which such subjects enjoy immediate justification for the target moral claims as the result of experiencing moral emotions.

In their studies of extraneously induced disgust, Wheatley and Haidt (2005) found that, when prompted in follow-up interviews, participants acknowledged that the extraneous factors in question were irrelevant to the truth of the moral claims at issue and agreed that they were not reliable indicators of moral truth. It is implausible that the participants think that the presence of the factors provides evidence for the moral claims in question. Thus, in cases of extraneously induced emotions, it is

plausible that participants do not rely on an inference from claims about the irrelevant factors to a moral claim. The data suggest instead that the morally irrelevant factors cause the participants to feel emotions, which the participants then use as (at least part of) the grounds for their moral attitudes. That the participants' moral judgments were formed on the (at least partial) basis of an emotion suggests that the attitudes in question may sometimes not be formed on the basis of an inference from the factors that influence the emotion.

So far we have evidence that subjects sometimes do not infer moral claims from non-moral factors, but instead (at least partly) ground their moral attitudes on emotions in cases of extraneously induced emotion. It is plausible that subjects ground their moral attitudes on emotions in a similar way when the emotions are not extraneously induced. There is no obvious reason to suppose that radically different psychological processes produce emotions from irrelevant factors than the ones that produce emotions from relevant factors. So, even in some cases where emotions are induced by morally relevant factors (e.g. Harman's Cat example) subjects may ground their moral attitudes on their emotional states rather than inferring the relevant moral claims from non-moral claims about the scenario. That is, on an empirically plausible account of subjects' moral belief forming processes in Harman's Cat example subjects do not infer the wrongness of the hoodlums' actions from claims about the non-moral properties of their actions. Rather, subjects experience moral indignation at the

hoodlums' act, and their moral indignation forms the basis of a justified moral belief.[90]

7 Conclusion

In this paper, I have responded to a recent empirical challenge—the Argument from Irrelevant Factors (AIF)—that threatens to undermine the widespread view that we have intuitive moral knowledge, in the psychologists' sense. I have defended the claim that some of the moral judgments we form suddenly and effortlessly, without relying on premises to which we have conscious access, constitute moral knowledge. The defense required a detailed analysis of the empirical data in question. The analysis revealed that the results do not in fact impugn the reliability of intuitive moral judgments.

The present argument advances the debate concerning the epistemic significance of the covert influence of morally irrelevant factors on moral judgment. Respondents to the AIF have so far conceded that the empirical results show at least that intuitive moral judgment is unreliable in the experimental contexts where the irrelevant influences have their effect. As I explained above, this response is unpromising as a defense of intuitive moral knowledge, because the irrelevant

[90] Further evidence that subjects do not arrive at moral attitudes as the result of inferences from non-moral claims about the scenario comes from cases of "moral dumbfounding" discussed in Haidt (2001). In such cases, participants feel intense emotional reactions to various claims—e.g. that incest between two consenting, sterile siblings is wrong—and continue to affirm the relevant moral claim, despite having their arguments in support of the claim undermined. A plausible account of the data is that the subjects ground their moral judgment on the wrongness of the act on their emotion (which does not dissipate in response to the counterarguments they encounter) rather than on an inference

influences in question are widespread.

My argument is not concessive, and it shows that the empirical findings most often cited in support of the AIF do not in fact show that intuitive moral judgments are unreliable, even in the experimental contexts in question. The argument requires distinguishing between fine-grained and coarse-grained moral judgments. While the results do suggest that fine-grained moral judgments (e.g., her act is wrong to degree 78/100) would be unreliable if made, this is unsurprising. And, in addition, it is not plausible that we regularly make such precise judgments. By contrast, the experimental findings do not suggest that coarse-grained moral judgments are unreliable. The effect sizes in results concerning the influence of morally irrelevant factors are too small to provide evidence that our coarse-grained moral judgments often differ in response to morally irrelevant factors.

In the final sections, I argued that various empirical results provide evidence that intuitive moral judgments are sometimes immediately justified by moral emotions. By arguing that the data that were supposed to undermine moral intuitionism can in fact be used to support it, I aimed to turn the tables on proponents of the AIF. I also hope to have helped to reframe the relationship between moral intuitionism and empirical psychology. In contrast to the recent trend to see them as opposed to each other, I have offered reasons to think that moral

---

from the sort of non-moral features of the scenario which the participants point to as they try to respond to counterarguments (and which the scenarios are carefully constructed to omit).

intuitionists should embrace and utilize findings in empirical moral psychology.

REFERENCES

Allman, J. & Woodward, J. (2008). What are moral intuitions and why should we care about them? A neurobiological perspective. *Philosophical Issues, 18*(1), 164-185.

Armstrong, D. M. (1973). *Belief, truth and knowledge*. Cambridge: Cambridge University Press.

Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, *21*(1), 147-152.

Bastick, T. (1962) *Intuition: How we think and act*. Chichester, England: Wiley.

Batty, C. (2010a). Olfactory experience I: The content of olfactory experience. *Blackwell Philosophy Compass*, *5*(12), 1137-1146.

Batty, C. (2010b). Olfactory experience II: Objects and properties. *Blackwell Philosophy Compass, 5*(12), 1147-1156.

Bengson, J. (2010). The intellectual given. Doctoral dissertation, University of Texas at Austin. Available electronically from http : / /hdl .handle .net /2152 /ETD -UT -2010 -05 -1367 .

Bengson (forthcoming). Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research*.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs 37*(4), 293-329.

BonJour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, *5*(1), 53-73.

Brady, M. S. (2007). Recalcitrant emotions and visual illusions. *American Philosophical Quarterly*, *44*(3), 273-284.

Brady, M. S. (2009). The irrationality of recalcitrant emotions. *Philosophical Studies*, *145*(3), 413-430.

Bruner, J. S. (1960) *The process of education.* Cambridge, Mass.: Harvard UP.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology*, *42*(1), 33-44.

Bruner, J. S., & Rodrigues, J. S. (1953). Some determinants of apparent size. *The Journal of Abnormal and Social Psychology*, *48*(1), 17.

Carter, L. F., & Schooler, K. (1949). Value, need, and other factors in perception. *Psychological Review*, *56*(4), 200-207.

Chudnoff, E. (2011). The nature of intuitive justification. *Philosophical Studies*, *153*(2), 313-333.

Clarke, S. G. (1986) Emotions: Rationality without cognitivism. *Dialogue*, *25*(4), 663-674.

Cohen, S. (1984). Justification and truth. *Philosophical Studies*, *46*(3), 279–295.

Comesana, J. (2006). A well-founded solution to the generality problem. *Philosophical Studies*, *129*(1), 27-47.

Conee, E. and Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies*, *89*(1), 1-29.

Cushman, F., Knobe, J., Sinnott-Armstrong, W. (forthcoming). Moral appraisals affect doing/allowing judgments. *Cognition.*

Damasio, A. (1994). *Descartes error: Emotion, reason, and the human brain*, New York: G.P. Putnam's Sons.

Deonna, J., & Teroni, F. (2012) *The emotions: A philosophical introduction*. Routledge. de

Sousa, R. (1987) *The rationality of emotion*, Cambridge, MA: MIT Press.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation*, *19*, 207-283.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*, 384-384.

Ekman, P. (2007) *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Holt Paperbacks.

Ekman, P., & Friesen, W. V. (1971) Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, *10*(2): 159-168.

Eskine, K., Kacinik, N., & Prinz, J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science, 22*(3), 295-299.

Feldman, R. (1985). Reliability and justification. *Monist*, *68*, 159–174.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, *51*(1), 23-43.

Fodor, J. (1988). A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'. *Philosophy of Science*, *55*(2), 188-198.

Foot, P. (1978).The problem of abortion and the doctrine of the double effect. In *Virtues and vices*. Oxford: Basil Blackwell.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, *73*(20), 771–791.

Goldman, A. I. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge*, Dordrecht: Reidel.

Goldman, A. I. (1986) *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. I. (1992). *Liaisons: Philosophy meets the cognitive and social sciences.* Cambridge, MA: MIT Press.

Goldman, A. I. (2008a). Immediate justification and process reliabilism. In Q. Smith, (Ed.), *Epistemology: New essays*, Oxford: Oxford University Press.

Goldman, A. I. (2008b). Reliabilism. In the *Stanford Encyclopedia of Philosophy http://plato.stanford.edu/entries/reliabilism/*

Goodman, N. (1978). *Ways of worldmaking*. Indianapolis: Hackett Publishing Company.

Greene, J. (2007). Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences. 11*(8), 322-323.

Greene, J. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, *45*(3), 581-584.

Greene, J., Sommerville, B., Nystrom, L., Darley, & J., Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293,* 2105-2108.

Greenspan, P. (1988). *Emotions and reasons: An inquiry into emotional justification.* London: Routledge.

Griffiths, P. E. (1998). *What emotions really are: The problem of psychological categories*. The University of Chicago Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Hansen, T., Olkkonen, M., Walter, S. & Gegenfurtner, K.R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367-1368.

Harman, G. & Thomson J.J., (1996). *Moral relativism and moral objectivity,* Wiley-Blackwell.

Harman, G. (1977). *The nature of morality: An introduction to ethics* Oxford University Press.

Hawthorne, J., & Lasonen-Aarnio, M. (2009). Knowledge and objective chance. In P Greenough, D. Pritchard & T. Williamson (Eds.), *Williamson on knowledge* (pp. 92-108). Oxford University Press.

Heck, R. G, Jr. (2000). Nonconceptual content and the 'space of reasons'. *The Philosophical Review*, *109*(4), 483-523.

Heller, M. (1995). The simple solution to the generality problem. *Noûs*, *29*, 501–515.

Helm, B. W. (2007). *Emotional reason: Deliberation, motivation, and the nature of value.* Cambridge University Press.

Hohmann, G. W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology*, *3*, 143-156.

Huemer, M. (2001). *Skepticism and the Veil of Perception.* Lanham, MD: Rowman & Littlefield.

Huemer, M. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research, 74*(1): 30-55.

Hutcherson & Gross (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, *100*(4), 719-737.

Jackson, A. (2011). Appearances, rationality, and justified belief. *Philosophy and Phenomenological Research*, *82*(3), 564-593.

James, W. (1884). What is an emotion? *Mind*, *9*, 188-205.

James, W. (1890). *Principles of Psychology*. New York: Holt.

Johnston, M. (2001). The authority of affect. *Philosophy and Phenomenological Research* *63*(1), 181-214.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissible harm*. New York: Oxford University Press.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(279), 190-194.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*(4), 315-329.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, Illinois: The University of Chicago Press.

Lange, C. (1885). The emotions. In E. Dunlap (Ed.) *The emotions*. Baltimore, MD: Williams & Wilkins (current edition 1922).

LeDoux, J. (1996). *The emotional brain: The mysterious underpinning of our emotional life*. New York: Simon and Schuster.

Lee & Schwarz (2011). Wiping the slate clean: psychological consequences of physical cleansing. *Current Directions in Psychological Science*, *20*(5), 307-311.

Lehrer, K. (1990). *Theory of knowledge*. Boulder, CO: Westview.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*(4), 501-512.

Levy, N. (2006). Cognitive scientific challenges to morality. *Philosophical Psychology, 19*(5), 567-587.

Liao, S. M. (2010). A defense of intuitions. *Philosophical Studies.* 140(2), 247-262.

Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues*, *21*(1): 289-311.

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.

Markie, P. (2005). The mystery of direct perceptual justification. *Philosophical Studies*, *126*(3), 347-373.

Markie, P. (2006). Epistemically appropriate perceptual belief. *Nous, 40*(1), 118-142.

McBrayer, J. (forthcoming) A limited defense of moral perception. *Philosophical Studies.*

McDowell, J. (1998). *Mind, value, and reality*. Cambridge: Harvard University Press.

McGrath, M. (forthcoming-a). Phenomenal conservatism and cognitive penetration: The 'bad basis' counterexamples. In C. Tucker (ed.) *Seemings and Justification*. Oxford University Press.

McGrath, M. (forthcoming-b). Siegel and the epistemic impact for epistemological internalism. *Philosophical Studies.*

McGrath, S. (2004). Moral knowledge by perception. *Philosophical Perspectives*, *18*(1),

209-228.

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, *45*(3), 577-580.

Nadelhoffer, T. & Feltz, A. (2008). The actor–observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. *Neuroethics*, *1*(2), 133-144.

Neta, R. (2010). Liberalism and conservatism in the epistemology of perceptual belief. *Australasian Journal of Philosophy, 88*(4), 685-705.

Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions.* Cambridge University Press.

Oatley, K, Keltner, D, & Jenkins, J. M. (2006). *Understanding emotions*. Oxford: Blackwell.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5), 1-16.

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald, S., & Munafò, M. R. (forthcoming). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological science.*

Petrinovich, L., and O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*(3), 145-171.

Pitcher, G. (1965). Emotion. *Mind*, *74*(295), 326-346.

Pizarro, D. & Bloom, P. (2003). The intelligence of the emotions: Comment on Haidt (2001). *Psychological Review*, *110*(1), 193–196.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion.* Oxford University Press.

Prinz, J. (2006). Is the Mind Really Modular? In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22-36). Oxford: Blackwell.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs*, *34*(4), 517-549.

Pryor, J. (2004). What's wrong with Moore's argument? *Philosophical Issues*, *14*(1), 349-378.

Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*(1), 111-132.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception?. *Behavioral and Brain Sciences*, *22*(3), 341-423.

Roberts, R. (2003). *Emotion: An essay in aid of moral psychology.* Cambridge University Press.

Schacter, S. & Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, *69*, 379-399.

Schnall, S., Haidt, J., Clore, G., & Jordan (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*(8), 2096-1109.

Shafer-Landau, R. (2003). *Moral realism: A defense*. Oxford University Press.

Shafer-Landau, R. (2008). In defense of moral intuitionism. In W. Sinnott-Armstrong (Ed.) *Moral psychology, volume 2: The cognitive science of morality: intuition and diversity* (pp. 83-96). Cambridge, MA: MIT Press.

Siegel, S. (2006). Which properties are represented in perception?. In T. Szabo Gendlre & J. Hawthorne (Eds.), *Perceptual experience* (pp. 481-503). Oxford University Press.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Nous*, *46*(2), 201-222.

Siegel, S. (forthcoming-a). The epistemic impact of the etiology of experience. *Philosophical Studies.*

Siegel S. (forthcoming-b). Replies to Fumerton, Huemer, and McGrath. *Philosophical Studies.*

Silins, N. (2008). Basic justification and the Moorean response to the skeptic. In T. Szabo Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology: Volume 2* (pp. 102-142). Oxford University Press.

Sinnott-Armstrong, W. (2006). Moral intuitionism meets empirical psychology. In T. Horgan & M. Timmons (Eds.), *Metaethics after moore*. Oxford, UK: Oxford University Press.

Sinnott-Armstrong, W. (2008). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.) *Moral psychology, volume 2: The cognitive science of morality: intuition and diversity* (pp. 47-76). Cambridge, MA: MIT Press.

Sinnott-Armstrong, W. (2011). Emotion and reliability in moral psychology. *Emotion Review, 3*(3), 288.

Solomon, R. (1976). *The passions: The myth and nature of human emotions*. New York: Anchor Press, Doubleday.

Sosa, E. (2007). *A virtue epistemology*, Oxford: Oxford University Press.

Stemmler, D. G. (1989). The autonomic differentiation of emotions revisited: convergent and discriminant validation. *Psychophysiology*, *26*, 617-632.

Stokes, D. (2012). Perceiving and desiring: A New Look at the cognitive penetrability of experience. *Philosophical Studies 158*(3), 477-92.

Stokes, D. (forthcoming). Cognitive penetrability. *Philosophy Compass.*

Street, S. (2006). A darwinian dilemma for realist theories of value. *Philosophical Studies*, *127*(1), 109-166.

Sturgeon, N. (2002). Ethical intuitionism and ethical naturalism. In P. Stratton-Lake (Ed.) *Ethical intuitionism: Re-evaluations*, (pp. 184-211). Oxford: Clarendon Press.

Tajfel, H. (1957). Value and the perceptual judgment of magnitude. *Psychological Review*, *64*(3), 192-204.

Tappolet, C. (2006). Emotions, perceptions, and emotional illusions. In Clotilde Calabi (Ed.), *The crooked oar, the moon's size and the Kanizsa triangle: Essays on perceptual illusions*. Cambridge, MA: MIT Press.

Taylor, G. (1975). Justifying the emotions. *Mind*, *84*(335), 390-340.

Thomson, J.J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*, 204-17.

Thomson, J. J. (1985). Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. *Yale Law Journal*, *94*, 1395-1415.

Tolhurst, W. (2008). Moral intuitions framed. In W. Sinnott-Armstrong (Ed.) *Moral psychology, volume 2: The cognitive science of morality: intuition and diversity* (pp. 77-82). Cambridge, MA: MIT Press.

Tucker, C. (2011). Why open-minded people should endorse dogmatism. *Philosophical Perspectives, 24*(1), 529-545.

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17,* 476-477.

Vayrynen, P. (2006). Some good and bad news for ethical intuitionism. *The Philosophical Quarterly, 58*(232), 489-511.

Watkins, M., & Jolley, K. D. (2002). Pollyanna realism: Moral perception and moral properties. *Australasian Journal of Philosophy*, *80*(1), 75-85.

Wheatley, T. & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780-784.

Witzel et al. (2011). Object knowledge modulates colour appearance. *i-Perception*, *2*(1), 13-49.

Wright, C. (2007). The perils of dogmatism. In S. Nucatelli & G. Seay (Eds.), *Themes from G. E. Moore: New essays in epistemology and ethics* (pp. 25–48). Oxford University Press.

Zhong, C. & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*(5792), 1451-1452.