

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XIX · NUMBER 1

January 2021

ARTICLES

Games and the Good Life

Michael Ridge

Objectified Women and Fetishized Objects

Paula Keller

Blaming for Unreasonableness:

Accountability without Ill Will

Alisabeth Ayars

The Moral Closure Argument

Matt Lutz

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XIX · NUMBER 1

January 2021

ARTICLES

- 1 Games and the Good Life
Michael Ridge
- 27 Objectified Women and Fetishized Objects
Paula Keller
- 56 Blaming for Unreasonableness:
Accountability without Ill Will
Alisabeth Ayars
- 80 The Moral Closure Argument
Matt Lutz

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Editor

Mark Schroeder

Associate Editors

Dale Dorsey

James Dreier

Julia Driver

Anca Gheaus

Errol Lord

Tristram McPherson

Colleen Murphy

Hille Paakkunainen

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

Elizabeth Anderson	Philip Pettit
David Brink	Gerald Postema
John Broome	Joseph Raz
Joshua Cohen	Henry Richardson
Jonathan Dancy	Thomas M. Scanlon
John Finnis	Tamar Schapiro
John Gardner	David Schmidtz
Leslie Green	Russ Shafer-Landau
Karen Jones	Tommie Shelby
Frances Kamm	Sarah Stroud
Will Kymlicka	Valerie Tiberius
Matthew Liao	Peter Vallentyne
Kasper Lippert-Rasmussen	Gary Watson
Elinor Mason	Kit Wellman
Stephen Perry	Susan Wolf

Managing Editor

Jesse Wilson

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

GAMES AND THE GOOD LIFE

Michael Ridge

AHUMAN LIFE devoid of play would be a deeply impoverished one. Play in childhood is especially important, but for human beings play remains an important ingredient in the good life throughout adulthood. Stuart Brown goes so far as to claim that a “life without play is a grinding, mechanical existence organized around doing the things necessary for survival.”¹ While equally strong claims about how essential playing *games* is to a good life are less plausible, it is remarkable just how many people *do* find playing games immensely rewarding. Indeed, a disposition not only to play but to play games is cross-culturally robust. Although games are perhaps “not for everyone,” it remains plausible that for a large portion of humanity playing games contributes to the good life. One might naturally wonder *how* does playing games so contribute? Granted, games can be very good, but what exactly is so good about them when they are good? Although a natural starting point, this question is perhaps naive. Games come in all shapes and sizes, and different games are often good in very different ways. Chess, bridge, bingo, Chutes and Ladders, football, spin the bottle, Dungeons & Dragons, Pac-Man, Minecraft, and charades can all contribute to a good life, but each will characteristically enrich life in its own distinctive way. Some games facilitate socializing, others improve physical fitness, others promote a sense of fairness and reciprocity, while others enhance concentration and analytic skills. Asking, “What is good about games?” with the presupposition that there is a simple, unified answer is as naive as asking (in the same spirit), “What is good about fiction?” “What is good about art?” or “What is good about sex?”

However, a less naive question in the vicinity is not hard to formulate. Plausibly, much of the heterogeneity of the value of games stems from the different *instrumental* value of different games. Perhaps we should therefore ask in what ways the *activity* of playing games is characteristically good *for its own sake*. Even here there may be heterogeneity. One and the same kind of activity can be good for its own sake in different ways in different contexts. In one context, dancing might be good for its own sake in virtue of its expressive value, as in the intimate

¹ Brown, *Play*, 11.

dance of two lovers, while in another context dancing might be good for its own sake in virtue of its aesthetic value, as when a professional dancer performs at the top of her game. Moreover, even a single token action can be good for its own sake in different respects. An act of kindness might be both morally good for its own sake and aesthetically good for its own sake if performed with grace and style. Even given these complications, it seems that the heterogeneity of the noninstrumental value of playing games is likely to be far more restricted and theoretically manageable than the heterogeneity of the value of playing games *tout court*. It is not crazy to suppose that playing games as such provides a reasonably restricted set of characteristic noninstrumental goods.

The two most striking candidate noninstrumental values characteristically associated with playing games are *play* and *achievement*. Here we need to distinguish “play” as used in “play a game” from “play” as used to refer to what I shall call “play (full stop).” As used in “play a game,” “play” is a transitive verb that gets its content in part from its direct object (compare “play a game” with “play the piano” and “play a joke on someone”). By contrast, one can play (full stop) without playing a game or indeed playing anything else. A child frolicking on a hill is playing (full stop), but need not be playing a game. More controversially, one can play a game without playing (full stop), as in the case of a jaded footballer who plays the game just for the money. Still, it is plausible that there is a non-accidental relationship between playing games and playing (full stop). The noninstrumental value of playing (full stop) is thus a reasonable candidate for explaining the noninstrumental value of playing games. Moreover, on many influential theories of welfare, play (full stop) contributes to the good life in and of itself.² These theories typically pick up on the idea that welfare is plausibly understood in terms of nature fulfillment, and argue that play (full stop) is an essential element of human nature.

Achievement is also a plausible candidate for explaining the value of playing games, though. When done well, playing a game can constitute achievement, and achievements are also often plausibly held to contribute to the good life in and of themselves.³ Moreover, we admire competitors for their achievements, and this admiration does not seem out of place. Admittedly, both these values, play (full stop) and achievement, can be found in non-game contexts. Even so, providing these goods may somehow be characteristic of why we value games, and these values may take on a special form when found in the context of games. To be clear, these candidates are not mutually exclusive, and a pluralist view of

2 Cf. Finnis, *Natural Law and Natural Rights*; Nussbaum, “Human Functioning and Social Justice.”

3 Cf. Griffin, *Well-Being*; Bradford, *Achievement*.

the value of games that includes both values has much to recommend itself; indeed, I defend a form of this view here.

Unfortunately, the philosophical literature on the noninstrumental value of playing games is sparse. One of the few sustained treatments of the topic can be found in an underappreciated exchange between Thomas Hurka and John Tasioulas.⁴ Interestingly, despite taking different views of what it is to play a game, they both make room for the noninstrumental value of play and achievement in game play *and* they both argue that these two goods stand in an important explanatory relation to one another. However, they take diametrically opposed views as to which of these good is more basic. Roughly, on Hurka's view, the good of achievement is more basic. The noninstrumental value of achievement explains the value of what Hurka calls "playing in a game."⁵ The idea is that if something is noninstrumentally good then loving that thing is also noninstrumentally good, and that playing in a game involves loving a noninstrumentally good activity for its own sake. In this way, the value of achievement in a game *grounds* the value of playing in a game. Tasioulas takes exactly the opposite approach. He argues that there must be something independently good about playing a game that grounds the value of achievement in that game. On his view, the typical grounding good or "framing value" of games is play itself—what I am here calling "playing (full stop)."

In this essay, I argue that while both contain important insights, neither Hurka's "achievement first" order of explanation nor Tasioulas's "play first" order of explanation is fully correct. I argue instead for what I call a "variable priority" view. On the view defended here, the value of play sometimes grounds the value of achievement in a game, while in other cases the independently grounded value of achievement in a game provides further grounding for the value of play. In the latter case we have only a *further* grounding because play is plausibly good for its own sake independent of any framing value in a way that achievement argu-

4 Hurka "Games and the Good"; and Tasioulas, "Games and the Good."

5 Hurka frames his view in terms of "intrinsic value," but there is room for a distinction between noninstrumental value and intrinsic value, as Korsgaard has famously argued and Hurka himself discusses elsewhere. See Hurka, "Two Kinds of Organic Unity"; and Korsgaard, "Two Distinctions in Goodness." See also Langton, who gives the nice example of the value of a wedding ring ("Objective and Unconditioned Value"). I might value my wedding ring in and of itself, and not as a means to some further good, but it is valuable only because of one of its relational properties—its history and role in my life with my spouse. An exact replica without that history would not have the same kind of value. So we can have noninstrumental value that depends on more than the intrinsic properties of the object of evaluation, which on one way of defining "intrinsic value" means we can have noninstrumental but nonintrinsic value. For my purposes it will be sufficient to focus on the noninstrumental value of games. Whether the value is also intrinsic is a further question I do not discuss here.

ably is not. Still, this further grounding explains why play in some game contexts is good for its own sake in a distinctive way—it explains why it is good in itself in a way that not all forms of play are.

I

Hurka's explanation of the noninstrumental value of playing games draws heavily on Bernard Suits's classic definition of "play a game." On Suits's view, playing a game has three elements: (1) the prelusory goal, which is a goal that can be understood independently of playing the game, (2) the constitutive rules, and (3) the lusory attitude, which is a matter of accepting the constitutive rules because of the activity they make possible. To play a game is to adopt the lusory attitude and seek the prelusory goal, but only in ways permitted by the constitutive rules. The easiest way to see how the view works is through an example. To play golf is to seek the prelusory goal (getting the ball into the cup), but only in ways compatible with the constitutive rules (one must use golf clubs as they are intended to be used to get the ball into the cup, rather than just dropping the ball in the cup by hand), where one accepts the rules just because they make this sort of activity possible. The constitutive rules are understood as ruling out some of the more efficient ways of achieving the prelusory goal, thus making it more of a challenge to achieve that goal. This account also allows us to make sense of the lusory goal, which is simply the prelusory goal achieved in ways permitted by the constitutive rules.

Hurka suggests that this analysis of "play a game" naturally suggests a view about what makes a game good. Games are designed to present us with challenges, so they must be difficult enough to challenge us. Thus, Chutes and Ladders and rock, paper, scissors are poor games. At the same time, they must not be so difficult as to be virtually impossible. Hurka's account of the value of playing games applies only to good games—ones that strike a happy medium between these extremes. Hurka argues that playing good games provides opportunities for achievement. On Hurka's account, achievements are activities in which one deliberately achieves a goal by overcoming some difficulty, and the greater the difficulty, the greater the achievement. Therefore, excellence in (good) games entails achievement. Hurka agrees with those philosophers who hold that achievement is a noninstrumental value, and so concludes that excellence in games is noninstrumentally good in virtue of constituting achievement. On this account, well-designed games are made to provide opportunities for achievement. As Gwen Bradford puts it in her landmark discussion, "Games are special

in that their very structure has something very close to the central elements of achievement built right in.”⁶

Achievement does not exhaust the noninstrumental value of game play on Hurka’s account. Hurka draws a distinction between playing a game and “playing in a game.” Even a jaded professional, who plays baseball just to get a salary, in one clear sense still counts as playing the game. Such a “pure professional” still counts as playing on Suits’s canonical definition because he adopts the rules to engage in that very activity—he just happens to have some further reasons for wanting to engage in that activity. Suits himself makes it clear that his use of the phrase “just because” is not meant to entail that this is the player’s only reason for accepting the rules, but that it always must be among the reasons. Hurka notes that when Suits finally defends the value of playing games he implicitly shifts to an understanding of game playing on which one must accept the rules *only* for the sake of the activity this makes possible. Suits’s own argument, which I discuss further in the final section of this paper, invites the reader to imagine a kind of “Utopia” in which all our instrumental goods are provided for. He argues that playing games would be the only activity we would find worthwhile in such circumstances, and that this tells us something important about the value of playing games. Hurka’s point is that those who play games in Utopia must accept the constitutive rules of the games they play *entirely* for the sake of engaging in the activity itself. At least, insofar as they are rational they must accept the rules for this reason, since there could be no instrumental reasons for accepting them. Players in Utopia play their games with an amateur attitude, and this is what Hurka calls “playing in a game.” This counts as playing (full stop) on Hurka’s view, because to play just is to engage in an activity for its own sake.⁷

Hurka argues that playing in a game is noninstrumentally good, but this further good is *derived from* the noninstrumental good of achievement. Here Hurka draws on his other work in normative theory, where he argues that if something is itself noninstrumentally good then the positive attitude of loving that thing for

6 Bradford, *Achievement*, 183.

7 Hurka, “Games and the Good,” 227. Hurka cites Suits as endorsing this view of what it is to play (full stop). Actually, Suits makes it clear that this is only “play” in an entirely stipulative sense (Suits, *The Grasshopper*, 161). On his considered view of “play” as we understand that notion in ordinary language, we must add an additional necessary condition—to count as playing (full stop) one must not merely do something for its own sake; this must also involve reallocating resources normally devoted to instrumental activities to the activity in question (Suits, “Words on Play”). Since those in Utopia have no instrumental needs, this entails that there can be no play in Utopia, which is in my view a highly counterintuitive upshot of Suits’s considered view of play (full stop). However, I put this complication to one side here as I ultimately defend a different conception of play (full stop) anyway.

the property that makes it good is also noninstrumentally good. If, for example, happiness is good for its own sake, then loving happiness (that is, benevolence) is also good for its own sake. Having argued that excellence in playing games is independently noninstrumentally good, he argues that because playing in a game involves loving the activity because of its good-making property (the difficulty of the activity), playing in a game is also noninstrumentally good. The value of play in a game is on this account an entirely *derived* good—its value is derived from the value of achievement. Playing in a game is thus itself not a fundamental good, but is derived from two more basic goods in a specific way. He suggests, though, that the value of playing in a game can be paradigmatic despite not being fundamental, “because it gives the clearest possible expression of a certain type of value.” His point here is that, in many other cases, achievement involves some further instrumental good, and this can distract us from the essential good of achievement as such. Because games take otherwise worthless ends and derive value from them by making them help constitute achievements, the value of achievement as such is more clearly on display in the case of games.

Hurka concludes by discussing evil achievements and achievements that involve some independently recognizable value, e.g., a moral value. He argues that the latter achievements are more valuable than achievements as in games in which the ends are in themselves worthless. This is why Nelson Mandela’s achievements are more valuable in themselves than Garry Kasparov’s.⁸ Still, achievements that involve morally neutral ends (like checkmate) can still be genuinely valuable achievements, just less so. In the case of evil achievements, as with someone who overcomes great difficulty to commit genocide, Hurka is at least willing to allow that such achievements lack noninstrumental value. The presence of an evil end is on this view a kind of “defeating condition” on the value of achievement.⁹

Tasioulas takes a very different view. He objects to Hurka’s reliance on Suits’s definition of “play a game,” which he argues is implausibly broad. In particular, he suggests that, on Suits’s view, the infliction of justified punishment and the waging of lawful war count, quite literally, as playing games, and that this is absurd. In the case of justified punishment, we have a prelusory goal of preventing criminal behavior. We then have constitutive rules that make this more difficult by forbidding the punishment of the innocent or punishing the guilty disproportionately. Those in the criminal justice system plausibly accept these rules just so they can engage in this kind of activity—that is, justified punishment of

8 Hurka, “Games and the Good,” 234.

9 Mentioned in Hurka, “Games and the Good,” 225n7.

offenders.¹⁰ He makes a similar point in the case of justified war, which aims at repelling an attack by another state (prelusive goal), but only in accordance with humanitarian laws (constitutive rules) that make the achievement of this goal less efficient.

Tasioulas allows that playing games is sometimes good because it involves achievement, but argues against Hurka's view that achievement provides the most fundamental and distinctive good associated with playing games. He points out that many games are simply not well understood in terms of striving or achievement. Games of chance, for example, are more about "the thrill of surrendering to fate and delighting in good fortune."¹¹ Children playing blind-man's bluff, factory workers engaged in an impromptu football match during their lunch break, or an elderly pensioner enjoying a weekly game of bingo are all further examples of games whose value is not well understood in terms of achievement. Indeed, Tasioulas suggests that finding the value of playing games primarily in terms of achievement is *ideologically* suspect—a "sophisticated manifestation of a problematic trend in modern life ... the invasion of play by the rhetoric of achievement ... a defence of games in the spirit of the work ethics."¹²

Tasioulas suggests that the more fundamental and more general (noninstrumental) good associated with playing games is playing (full stop).¹³ He reminds us that playing is plausibly partly constitutive of welfare for creatures like us, drawing on both the work of philosophers analyzing welfare as well as documents like the UN Convention on the Rights of the Child, which includes a right to play.¹⁴ Insofar as playing a game characteristically involves playing (full stop), it is the playing that most fundamentally constitutes its noninstrumental value. Play (full stop) plausibly can be found in the kinds of games Tasioulas argued were not well understood in terms of striving for achievement. However, Tasioulas does not deny that games are sometimes good because they are vehicles for achievement. When they are valuable in this way, though, he argues that this is

10 Tasioulas, "Games and the Good," 238.

11 Tasioulas, "Games and the Good," 241.

12 Tasioulas, "Games and the Good," 251.

13 Actually, Tasioulas himself does not sharply distinguish playing a game from playing (full stop), perhaps because he would deny that there is a sharp distinction to be drawn there. He may think that all play is in some sense playing a game and that all game play involves what I would call "play (full stop)." Insofar as one allows, as I argue we should, that there is a distinction to be drawn here, though, it seems clear that what Tasioulas has in mind is what I am calling play (full stop). I therefore impose this distinction on the text in the interest of clarity in the present framework.

14 Tasioulas, "Games and the Good," 241.

typically itself explained by the value of play, thus inverting the order of explanation found in Hurka.

Tasioulas argues that Hurka's account of achievements is, like his (and Suits's) account of games, overbroad. He gives the example of someone ("Joe") whose dominant pursuit in life is counting blades of grass. He tries to count the exact number of blades of grass in as many lawns of a certain size as possible within a calendar year, remembering at the end of the year how many blades were in each year and also accurately describing each yard's general condition. This is the prelusory goal of his "game." The constitutive rule that makes his pursuit of this goal less efficient is a rule that requires that the counting is done manually and without assistance, that it must continue for twelve hours per day, and that on every other day he must fast. Joe adopts these rules entirely for the sake of the activity that they make possible. The game is Joe's own creation, and it affords him a sense of fulfillment but little pleasure. Joe becomes highly proficient at this "game" and can count and accurately recall ten times as many lawns per year as a hypothetical amateur. Tasioulas argues that this constitutes an achievement on Hurka's view (he does deliberately achieve a difficult goal), but that this is a reduction of Hurka's conception of achievement on the grounds that there is nothing valuable about what Joe does, apart from perhaps the incidental value of practical and theoretical reasoning in which he engages in the course of these activities.¹⁵

Tasioulas argues that what this example illustrates is that overcoming difficulty per se is not noninstrumentally valuable. On his view, overcoming difficulty constitutes an achievement, and is therefore valuable, only when the activity is independently valuable in some *other* way. In the case of playing games this "framing value" is "typically" play itself.¹⁶ The presence of some additional independent value might be understood as what Jonathan Dancy would call an "enabling condition" for the value of achievement.¹⁷ Tasioulas argues that if we modify his example of Joe and the game of counting blades of grass case so that it begins to look more and more like it exemplifies play (full stop) then it becomes more and more plausible to suppose that his excellence in this activity does indeed constitute an achievement. In particular, if we change the case so that Joe *does* enjoy the activity for its own sake, is regarded by him as "not serious," and is done only in his free time, filling no more than seven days per year and imposing no health-endangering burdens, then the idea that what he does can constitute an achievement is more plausible. If we further add that other people take part in the activity, so that it has a social dimension in Huizinga's sense, then it becomes

15 Tasioulas, "Games and the Good," 257–58.

16 Tasioulas, "Games and the Good," 243.

17 Dancy, *Ethics without Principles*; compare McKeever and Ridge, *Principled Ethics*.

even more plausible that what he does constitutes an achievement.¹⁸ If, for example, we add that others take part in the activity competitively and that they have an annual festival, then the idea that his prowess at counting blades of grass could ground an achievement becomes much more palatable.

We find two starkly different views of the noninstrumental value of playing games in Hurka and Tasioulas, and two different views of the nature of play and games themselves. Both agree that the noninstrumental value of playing games can involve both play (full stop) and achievement, but they differ profoundly on which of these two goods is more basic. This disagreement rests on a further disagreement about the nature of achievement and what makes it valuable. In the next section I argue that while each of these perspectives contains important insights, neither is quite right.

II

Although I do not think Suits's analysis of "play a game" is quite right, my differences with him do not matter for present purposes.¹⁹ Indeed, the argument for a "variable priority" view of the value of play (full stop) and achievement in playing games should go through on any otherwise plausible account of playing games. That being said, I should underscore that in my view Suits is onto something important and insightful in emphasizing the way in which games characteristically provide opportunities for achievement in virtue of being constituted by rules that make the achievement of some goal more difficult. Overcoming these difficulties can constitute achievements, and in my view providing opportunities for achievements is at least one of the main *functions* associated with the concept of a game. It is not, in my view, the only function associated with the concept of a game, and some of Tasioulas's examples (e.g., games of chance) suggest that there are other relevant functions—socializing or simply playing (full stop) and having fun may be other germane functions here. Though I prefer a more functionalist analysis of "play a game," which can make room for these heterogeneous functions associated with the concept of game play, my arguments here will not rely on this. On either my own view or Suits's view, games not only can provide opportunities for achievement, their doing so is an important characteristic of games, or at least of certain forms of game play.

I also agree with both Hurka and Tasioulas that a plausible account of the value of playing games will advert both to playing (full stop) *and* achievement. However, I must briefly register that I understand play (full stop) differently from

18 Huizinga, *Homo Ludens*.

19 Ridge, "How to Play Well with Others."

Hurka. Hurka follows Suits's stipulative definition of "play" and glosses play as doing something entirely for its own sake. For Hurka's purposes, this can be read as a harmless stipulation. However, my positive account of the noninstrumental value of playing games will advert to the value of play (full stop) *as that concept is understood in ordinary thought and discourse*, so for present purposes the contours of that concept matter. In particular, it is important to my account that play (full stop) is valuable in itself; this is how it can on my account ground the value of achievement. However, I do not think that merely doing something entirely for its own sake makes that activity itself good for its own sake. This is most clear in cases in which the activity is banal or evil.

Moreover, doing something for its own sake is neither necessary nor sufficient for playing (full stop) in the ordinary-language sense of "play," though Hurka is not alone in characterizing play in these terms. Indeed, the preponderance of accounts in the literature take it to be necessary.²⁰ For a start, very young children and simple-minded animals may not have the conceptual sophistication needed to do something for its own sake, yet we correctly judge that they play—in fact, young children provide paradigm instances of play. Moreover, doing something for its own sake is not sufficient for playing (full stop). Suits follows his arguments where they lead and suggests that a soldier who should be defending his city but instead engages in religious contemplation for its own sake is thereby playing.²¹ Intuitively, though, religious contemplation for its own sake is far too serious and contemplative to count as play. Indeed, religious adherents might well consider the characterization of such activities as mere play as sacrilege. The available empirical evidence suggests that being done for its own sake is *not* what guides ordinary speakers' judgments that some activity is play. In a study of the characteristics ordinary people use to guide their judgments of what counts as play, Smith and Vollstedt went so far as to conclude that "the association of Intrinsic Motivation with play was to be found *insignificant and negligible* in amount in this study, and *its status within any definition of play should be questioned seriously*."²²

An initially tempting hypothesis is that to play (full stop) just is to do something for the fun of it. However, play (full stop) requires *activity* in a sense that goes beyond merely "doing something." Riding a roller coaster is doing something, and typically something one does for the sheer fun of it, but it would be odd to call riding a roller coaster playing. Perhaps we should instead define "play" (full stop) as engaging in an activity for the fun of it, where "activity" is spelled

20 Hurka, Tasioulas, and Suits all take this approach; see also Burke, "'Work' and 'Play.'"

21 Suits, "Words on Play."

22 Smith and Vollstedt, "On Defining Play," 1048.

out in some way that rules out passive behaviors like riding a roller coaster. This is closer to correct, but it is still too broad. Intuitively, play (full stop) also involves an element of *spontaneity*. This is why someone rolling around in the mud, wrestling, or tickling someone are such clear cases of play. These activities are in an important sense *unscripted*. We should therefore, in my view, require that an agent's behavior is *unscripted* before we conclude that he is playing. These considerations suggest the following definition:

An agent *A* is playing if and only if (and by definition) *A* is engaged in an unscripted activity for the fun of it.

Of course, a full defense of this definition would require a separate paper. In what follows, I will implicitly rely on this definition to focus discussion. The details of this definition are *not*, however, essential to the main theses of this paper. All that is essential is the more basic idea that play (full stop) should *somehow* be understood in terms of doing something for the *fun* of it, rather than in terms of doing something for its own sake. To be clear, this is essential because the element of fun is in my view part of what explains why it is plausible to take play to be good for its own sake. This marks an important contrast with Hurka's more austere conception of play.

So much for "play a game" and "play (full stop)." Finally, how should we understand achievement and its value? Here I side with Hurka's account of the nature of achievement, but agree with Tasioulas about what makes achievement valuable when it is. On Hurka's account, achievement is constituted by deliberately achieving a goal in the face of challenges.²³ Tasioulas argues against this conception by appealing to our intuitions about the person who makes a game of counting blades of grass. His argument implicitly but crucially relies on the premise that something's being an achievement entails that it is valuable; this bridges the gap from "is worthless" to "is not an achievement." This implicit premise is dubious, though. Ordinary speakers would not hesitate to characterize an elaborately planned and flawlessly executed art heist as an achievement while in the same breath denying that it is good. Admittedly, "achievement" has the *connotation* that the activity in question is good in some way, but it would be implausible to build this connotation into the semantic content of "achievement."²⁴ In Gricean terms, this implication is cancellable. However, Tasioulas's

23 A much more detailed account in the same vein can be found in Bradford, *Achievement*. I return to Bradford's work below.

24 In fact, the case of evil achievements is far more complex than I can do justice to here. For useful discussion of evil achievements, see Bradford, "Evil Achievements and the Principle of Recursion." Bradford distinguishes the process and the product involved in an achieve-

evaluative thesis itself—that the counting of blades of grass in his original example is worthless in itself—seems plausible, and this is all he needs to undermine Hurka’s view. In the following section, I return to Tasioulas’s view of the need for a framing value for achievements to be valuable in themselves. For now, I will simply take this view as correct without further argument.

A plausible account of the noninstrumental value of playing games will ideally help explain why we have reason to play those games. The value in question should, in other words, somehow ground or reflect reasons that could and naturally would appear on the radar of an enthusiastic gamer. However, the value of achievement as such does not seem well suited to play this role. If you ask someone why they play baseball, chess, or cricket, you are unlikely to get the answer, “It is a good vehicle for achievements, and I want some achievements.” Someone who is passionate about chess, for example, will instead advert to features of the game itself—the elegance of the rules, the beauty of a stunning combination, or the rich history of the game. In some cases, they might advert to more abstract considerations, including the moral or symbolic value of the game. A chess player might, for example, emphasize the inherent fairness of chess—in chess, unlike “the real world,” consistently playing “good moves” will win, whereas “bullshit loses” (the contrast with, e.g., “moves” in a political debate is illustrative). Sometimes a move can be highly unexpected in a way that will strike an experienced player as funny, or even hilarious, so comedic reasons can also play a role. Purely subjective reasons can also figure in the reasons of a passionate gamer: “Playing basketball is just good fun!” The fact that some games provide an opportunity for friendly banter and other forms of socializing can also provide the right kind of reason. Finally, a whole host of instrumental reasons might figure in the reasons a passionate gamer might give for playing a given game—that it makes them more alert, more physically fit, more able to focus, better at strategic thinking in other contexts, and so on.

By contrast, “that it allows me to make great achievements” does not seem like the sort of reason a passionate gamer would naturally mention. While having some achievements plausibly contributes to having a good life, it does not seem

ment, and argues that the process always retains some positive value, but that this can be outweighed by the negative value of the product. She also argues that the process and product can together form a kind of organic unity in Moore’s sense. In the case that concerns me here, where the end is neither good nor evil but neutral (the state of affairs of a set of blades of grass having been counted), she would insist that the process is still valuable *qua* achievement. On her view this is an important point because achievement in games is typically like this—checkmate is in and of itself evaluatively neutral, e.g., as is getting a golf ball in a hole. Obviously if this view is right then premise 1 of Tasioulas’s account is also in trouble. I try to explain in the text why I find Tasioulas’s view more plausible.

to provide the “right kind of reason” for *first-person deliberation* about whether to play a given game or indeed whether to spend time playing games at all. This sort of consideration seems somehow at once both too abstract, not picking up on the specific features of the game the person loves, and too self-focused, almost narcissistic. Here some of Scanlon’s points about the role of welfare and of achieving one’s own ends in first-person deliberation seem very much to the point:

From an individual’s own perspective, taking these goals as not yet determined, we can say that a life goes better if the person is more successful in achieving his or her main rational goals (whatever these turn out to be), but the conception of well-being that can be formulated at this level is too indeterminate, too abstract, to be of great weight.²⁵

None of this is to downplay the normative significance of achievement. It is instead to help locate the perspective from which that significance is most salient. In my view, the consideration, “this will facilitate achievements for Smith” is most salient not for Smith himself, in deciding what course of action to pursue, but for those who care about Smith. Perhaps most obviously, if Smith is a child then Smith’s parents might choose activities for Smith in part because they will be good vehicles for achievement for Smith. Even if Smith is an adult, “that it will allow him to have some achievements” might also figure in how Smith’s friends think about how to advise and otherwise direct him. If, though, we are interested in getting clear on the noninstrumental value of playing games as it figures in the deliberation of the person playing the game, then a focus on achievement as such is misguided.

However, something in the vicinity of achievement is salient from a first-person perspective. Although an avid chess player would not characteristically say that they play the game for the sake of having achievements, they might well say that they play the game because it provides an engaging challenge. Competently overcoming challenges just is achievement (I am here assuming with Hurka and Bradford, anyway), so these reasons are obviously related, but the distinction is important. Most obviously, if your reason is “this will provide me with achievements” then if you fail to overcome the relevant challenges then this reason will not justify your action. If, however, your reason is simply, “this will provide an engaging challenge,” then so long as you engage with the challenge the reason can justify your action even if you fail to *overcome* those challenges.

Hurka’s discussion moves seamlessly between these two kinds of reasons. He often characterizes the relevant reasons in terms of achievement, and he begins his discussion by appealing to what makes us admire elite athletes. He

25 Scanlon, *What We Owe to Each Other*, 133.

therefore often focuses on the value of “*excellence in games*,” rather than simply playing games.²⁶ However, in other contexts, he appeals to difficulty rather than achievement to explain the value of what he calls “playing in a game,” where this is simply playing a game with the right attitude—out of love of it for its difficulty. Since it is possible to play in a game without manifesting excellence in the game, these are clearly different objects of evaluation. With this distinction in hand, I now return to Tasioulas’s view.

Tasioulas’s suggestion that play (full stop) is the typical framing value for achievements in a game derives some *prima facie* plausibility from my suggestion that what makes a practice a game is its function, and one of the relevant functions is play (full stop). However, this may make play in some ways an especially characteristic framing value for achievements, in a game, but it does not follow from this that the value of achievement in games is “typically dependent on the value of play itself.”²⁷ Even restricting our attention to other *noninstrumental* values that can be manifest in playing games, play is not the only candidate value. Other candidate noninstrumental values commonly found in game play include beauty, humor, and various forms of symbolic value. More importantly, it is not at all obvious why the *instrumental* value of playing games could not provide the needed framing value. It is the sheer pointlessness of counting blades of grass that makes achievement in that domain seem worthless. If we modify the case so that counting blades of grass has substantial instrumental value, then the idea that achievement in that domain is not valuable as such is much less powerful. Since most actual games provide all sorts of instrumental values for those who play them (physical fitness, mental acuity, self-esteem, a sense of fairness and reciprocity, etc.), these values could in turn provide the needed “framing” value to explain the value of achievement in those games. These values can, moreover, be present even when those playing a game are not thereby playing (full stop). The jaded footballer who plays just for a salary can still improve his physical fitness by playing, and can also produce beautiful plays within the game, for that matter, thus instantiating other instrumental and noninstrumental values.

Where does this leave us with regard to the contrasting orders of explanation found in Hurka’s and Tasioulas’s accounts? Recall that Hurka took the value of achievement as basic and explained the value of play in a game in terms of that value, whereas Tasioulas explained the value of achievement in a game in terms of the value of play (full stop) as a more basic “framing” value. I agree with Tasioulas at least this far: play (full stop) does sometimes provide the framing value for the value of achievement in a game. Sometimes, but not always. As I argued

26 Hurka, “Games and the Good,” 217, emphasis added.

27 Tasioulas, “Games and the Good,” 243.

above, one can play a game without playing (full stop), and such forms of game play plausibly can provide opportunities for valuable achievements. In these cases, play does not ground the value of achievement in a game, but (and here too I agree with Tasioulas and against Hurka) that value stands in need of *some* framing value. In my view, this framing value could take many forms depending on the game: it might be the beauty of the game, its symbolic value, or any of a wide range of instrumental values.

Furthermore, it can work the other way around: the value of play (full stop) can, in at least some cases, be partly explained by the value of achievement. Take a case in which the value of achievement is grounded, in part, in something other than play (full stop). Suppose someone wins a hard-fought chess game against a worthy opponent. This constitutes an achievement, and let us suppose it is valuable *qua* achievement, in part, because the winner found an elegant and beautiful combination. The beauty of this combination provides at least one of the framing values that explains why the achievement has noninstrumental value. We can allow that there will be other framing values, including the pleasure he takes in the game, for example. Crucially, once we allow that we have an achievement in virtue of the winner having done something difficult that is independently valuable in some way (beautiful and enjoyable), we plausibly have something that is noninstrumentally valuable *qua* achievement. In that case, if we agree with Hurka that it is noninstrumentally good to love something that is good for its good-making properties, then insofar as playing (full stop) in the game is partly constituted by loving the game for its difficulty and beauty, then playing (full stop) in the game will also be noninstrumentally valuable. In this sort of case, play (full stop) is plausibly noninstrumentally valuable in two ways. First, simply in virtue of being play, it is noninstrumentally valuable in virtue of partly constituting the person's welfare—or so many prominent theories of welfare contend (see references above). Second, insofar as play constitutively involves enjoyment, and in this case it is enjoyment taken in something that is itself good for its good-making properties, play is also noninstrumentally good in the same way that benevolence is noninstrumentally good—as an instance of loving the good for its good-making properties.

On my account, then, we therefore should take a “variable priority” view of the explanatory axiological relationships between play (full stop) and achievement in a game. Sometimes, the value of play provides a framing value that partly explains the value of achievement in a game (or at least explains why it is *so* valuable—see below). Sometimes, the value of achievement is instead partly explained by other values and the value of achievement in turn partly explains the value of play *qua* loving the good. Achievement, though, is itself what is good

about excellence in a game, as opposed to playing a game at all, and is not the sort of value that would typically be cited by someone passionate about a game as their reason for playing the game. Taking on an engaging challenge is, however, closely related to achievement as a value, since someone who takes on an engaging challenge is thereby trying to do something that, if they succeed, will constitute an achievement. Moreover, the value of taking on an engaging challenge is the sort of value that could plausibly figure among the reasons a player might have for playing a game. The challenge must be an engaging one for this to provide the agent's reason for playing the game because difficulty itself requires some sort of "framing value" to provide a noninstrumental value and hence a reason to play it.

My argument in this section has relied heavily on the idea that achievements require some framing value or other to be good for their own sake. However, more needs to be said about this admittedly controversial thesis. In the following section I provide the thesis with some additional defense and then explain how a modified version of the variable priority view could survive conceding this point.

III

In the previous section I relied on the idea that achievements require some framing value to be good for their own sake, and Tasioulas's example of counting blades of grass lends some credence to this view. However, perhaps I have been too swift in my rejection of the Hurka/Bradford view that no such framing values are essential. One might allow that achievements like those of the grass-blade counter have some value while admitting their value is minimal. Moreover, Guinness World Records is filled with achievements one might think have no framing value, yet many ordinary folks seem to admire people for these achievements.²⁸ To take one example, as I write this, the man who held the world record for the longest fingernails has just cut them off.²⁹ Plausibly the state of affairs of having fingernails over 900 cm long (roughly the length of a London bus) is as worthless as the state of affairs of having counted the blades of grass in one's front lawn. Yet, it seems, many people view such achievements as valuable in virtue of their manifestation of perseverance. We seem to admire the raw grit and stick-to-itiveness manifested in such achievements.

Gwen Bradford, in her landmark discussion of achievement, provides a deeper theoretical rationale for this view of the value of achievements with no framing value.³⁰ She begins with the perfectionist idea that the "excellent exercise" of

28 Thanks to an anonymous referee for pressing me on this point.

29 Stubbings, "Owner of World's Longest Nails Has Them Cut after Growing Them for 66 Years."

30 Bradford, *Achievement*.

certain distinctive human capacities are themselves valuable for their own sake. She then argues that in achievements we find two distinctively human capacities, namely rationality and the will, being excellently exercised. This is because, on her view, achievements involve the competent causation of some intended outcome in the face of difficulty, where competent causation requires the use of one's rationality and overcoming difficulty requires the exercise of one's will. The excellent exercise of each of these capacities is, given perfectionism, good for its own sake. Achievement, then, is the combination of the exercise of these two capacities. Bradford further argues that when we find unity in a diversity of valuable elements that we have a Moorean organic unity—a whole whose value is greater than the sum of the value of its parts. Achievements are, on her account, just such a case of unity in diversity: diversity because it involves two different capacities, but unity because they are coherently exercised in the same rational course of action. Achievements are therefore good because they constitute the unified exercise of two diverse capacities, and the exercise of each is itself independently valuable.³¹

Crucially, *nothing* in this explanation of the value of achievements requires that there be any “framing” value, and Bradford effectively endorses this view. However, Bradford's account of the *relative* value of achievements helps accommodate at least some of the intuitions behind those who, like myself and Tasioulas, think some framing value is necessary for an achievement to be good for its own sake. She argues for a version of a doctrine she takes from Hurka, the “*amare bonum bonus*,” which asserts that having a positive orientation toward something that is valuable is itself valuable for its own sake in virtue of being oriented toward the good. As applied to achievements, this means that when some framing value *is* present it makes the achievement itself more valuable for its own sake. This is because in achievement one pursues some end and “pursuit is a pro-activity” and so the kind of orientation that is needed for the application of the *amare bonum bonus*.³² On Bradford's view, while it is not true that a framing value is *necessary* for an achievement to have *any* value at all, the presence of such framing values can make such achievements much *more* valuable for their own sake. One might reasonably suppose that this is enough to accommodate what is insightful in the view that Tasioulas and I favor.

The dialectic at this point becomes subtle, since the difference between the two views might be one on which an achievement with no framing value, like counting blades of grass, has *no value* and one on which such achievements have

31 Bradford, *Achievement*, 126.

32 Bradford, *Achievement*, 165.

*very little value.*³³ A full discussion of the competing virtues of these views would require too much of a tangent for present purposes. Fortunately, those more sympathetic to Bradford's view can happily endorse something that is very close to the main thesis of this paper, the "variable priority view." I return to this issue below. In the meantime, I will very briefly try to summarize my reasons for preferring the view that some framing value is essential.

Start first with the more intuitive argument that framing values are not necessary for an achievement to be valuable for its own sake—the argument from cases. Consider again what we might call the "argument from Guinness World Records." To some extent, we may here simply have a clash of intuitions. However, it is not clear how wide the range of cases is in which this is a plausible diagnosis, given the wide variety of framing values available. Many of the achievements one finds in Guinness World Records do plausibly have suitable framing values. For example, many of these achievements involve athleticism, and so are typically at least instrumentally valuable. Being the "world's oldest bodybuilder" presumably helps prevent the loss of muscle mass that naturally accompanies aging, for example. Athletic achievements, and many of the other achievements one finds in Guinness World Records, are often enjoyable—think of the "runner's high" or the pleasure a bodybuilder takes in a "good pump," not to mention the exhilaration of some of the more adrenaline-inducing achievements. Other world records involve aesthetic value.

What about those achievements that do not involve any such framing value? Perhaps the world's longest fingernails provide a useful test case, particularly given the costs borne by the person who holds the record, who apparently can no longer open his left hand from a closed position and is by his own account, "in pain. With every heart beat all five fingers, my wrist, elbow and shoulder are hurting a lot and at the tip of the nail there is a burning sensation always."³⁴ Surely if we consider this achievement to be valuable, as its inclusion in Guinness World Records arguably suggests, then achievements with no framing value cannot only be valuable, they can be quite valuable—enough to warrant significant sacrifice. I must admit that, in this case, and in other cases with a similar structure, I do not share the intuition that the achievement in question is valuable.

However, I agree that there plausibly is something valuable "in the vicinity" of the achievement. After all, it does take enormous, if misguided, perseverance

33 Actually, it is unclear whether Bradford herself can insist that such achievements have very little value, given her characterization of the way in which more difficulty and more rationality as such enhance the value of any given achievement. Still, a view very much like Bradford's could take the line mentioned in the text.

34 Stubbings, "Owner of World's Longest Nails Has Them Cut after Growing Them for 66 Years."

to achieve something like this, and I agree that perseverance is a virtue—even when that perseverance is misdirected. Crucially, though, what follows from this is that the agent possesses something valuable—namely the virtue of perseverance. However, it is the virtue *itself*, and not its manifestation, that I consider valuable in itself. It is the agent's character that is valuable in itself, and that would have been valuable even if he had not had the opportunity to manifest his virtue. It simply does not follow from the fact that an activity manifests a virtue that the activity is itself valuable for its own sake—particularly once we reject the unity of the virtues and allow that a given virtue's manifestation in a particular context might be misdirected in this way.

In some passages, Bradford moves seamlessly from “displays a virtue” to “is valuable,” but this inference requires argument. This passage is a good example: “the most plausible account of why climbing a mountain has a value greater than zero is, perhaps, because it is a great display of virtue.”³⁵ Of course, there is room for reasonable disagreement over whether the manifestation of a virtue, in addition to the virtue itself, is valuable as such; nothing is obvious here. My point is simply that my slightly more austere view of the value of achievement can *agree* that achievements are a *sign* or *indicator* of something valuable (a virtue) even when there is no framing value in play so that the activity itself is not valuable. The conflation of the value of a virtue with the value of its manifestation arguably debunks the intuitions that might seem to favor the opposing view; it at least shows that more argument is needed.

Obviously much more could be said about the intuitive case for Bradford's view, but I can only devote so much space to this issue. What, though, about the more theoretical argument—her argument from perfectionism, as we might call it? In my view there is a step in Bradford's argument that requires further support but that seems to be taken as obvious. I have in mind the move from “achievement as such involves an exercise of the will” to “achievement as such involves an *excellent* exercise of the will.” It is only the latter that entails that achievement is valuable in light of the more general perfectionist framework as Bradford articulates it, but the former is all that is obvious from the proffered analysis of achievement. Here is a representative passage:

What is valuable is the *excellent exercise* of the perfectionist capacities. . . . Engaging in difficult activity requires the *exercise* of the will, which, as we have seen, is among the perfectionist capacities. It follows that difficult activity . . . is intrinsically valuable.³⁶

35 Bradford, *Achievement*, 87.

36 Bradford, *Achievement*, 121.

As stated, the argument is a *non sequitur*. Charitably, we can assume that the suppressed premise is that, when the activity is difficult, the exercise of the will is thereby excellent in the needed sense. This is hardly obvious, though. Critical here is just what makes a deployment of one's will count as "excellent" in the relevant sense. My own view is that it is only when the will is directed at some end or activity that is itself valuable in some way that the deployment of the will should count as excellent in the sense that would make the activity valuable *qua* perfectionism. A useful comparison here might be with Kant's conception of the value of the good will, though of course one need not endorse all of Kant's theoretical machinery to think that a use of one's willpower on something that is itself worthless or even evil should not count as an "excellent" deployment of the will.

These sparse arguments hardly settle the dispute between myself and Tasioulas on the one hand, and Hurka and Bradford on the other. However, even if Hurka and Bradford are right and Tasioulas is wrong, a modified version of the variable priority thesis is defensible. Suppose for the sake of argument that a view along the lines of Hurka's and Bradford's is correct. Even if achievements do not require a framing value to be valuable *at all*, both Hurka and Bradford allow that achievements can be much *better*—and better for their own sake—in virtue of the agent being motivated by some independent value of the activity. This is the *amare bonum* bonus principle. Although these are not framing values in the sense I have so far invoked, they play a similar theoretical role. Rather than explaining why a given activity is valuable at all, they explain why it is *so* valuable. Here, too, I would defend a sort of variable priority thesis of the relationship between play (full stop) and achievement. In some cases, the value of play explains why a given achievement is as valuable as it is—the play element explains why the activity is independently worthwhile and so explains why the *amare bonum* bonus principle applies. In other cases, though, the value of achievement is more basic, and the value of loving the activity (playing in a game, in Hurka's sense) is explained by the independent value of achievement. So even if my preferred view of the value of achievement, in general, as requiring some framing value is rejected in favor of the Hurka/Bradford view, a sort of variable priority thesis is still vindicated, albeit a weaker one.

IV

Suppose that the view of the noninstrumental value of playing games I have been sketching in the preceding sections is at least broadly on the right track. One might at this point reasonably wonder what, if anything, is especially *distinctive* about the value of playing games. After all, it is not as if any of the values

invoked in the preceding discussion cannot be found in other contexts. Playing a game well can be a vehicle for achievement, and for taking on an engaging challenge, but so too can scientific research, being a good parent, writing a novel, and a host of other kinds of activities. Playing a game can be a source of play (full stop), but one can play (full stop) without playing a game. Playing a game can be a source of beauty and fun, but these values can also quite obviously be found elsewhere. Perhaps we should conclude that, although playing games is an important source of value, it does not provide a distinctive kind of value that cannot be found elsewhere.

This is right as far as it goes, but we must be careful not to overlook other ways in which games might be distinctive sources of value. One way is to provide a value that cannot be found, or cannot easily be found, elsewhere, but that is not the only way. Games might instead be an especially *resilient* source of a given value, and this may itself be an important point. Indeed, I take this to be the main point of Bernard Suits's classic discussion of the role of games in Utopia. The groundwork for this point was laid in book 1 of *The Grasshopper*, where Suits (in the guise of the Grasshopper) argues that prudence is self-defeating:

Prudential actions (e.g., those actions we ordinarily call work) are self-defeating in principle. For prudence may be defined as the disposition 1/ to sacrifice something good (e.g., leisure) if and only if such sacrifice is necessary for obtaining something better (e.g., survival), and 2/ to reduce the number of good things requiring sacrifice—ideally, at least—to zero. The ideal of prudence, therefore, like the ideal of preventative medicine, is its own extinction.³⁷

Many of the activities we find worthwhile in daily life are done only out of necessity—for the sake of some greater good. Insofar as we could get those greater goods in some other way, these activities would be pointless. Suits's discussion of Utopia at the end of *The Grasshopper* puts this point into sharp relief by inviting the reader to imagine a world with no scarcity of any kind. Not only is there no scarcity of material resources, there is no scarcity of information or knowledge. Suits argues that many of the activities we find worthwhile in the actual world would seem pointless in Utopia. The production of material goods and scientific investigation, for example, would be pointless since by hypothesis all our material needs are already provided for and all knowledge has already obtained and can be had instantaneously from computers that store it. Suits argues that the playing of games would remain as the only intrinsically valuable activity in Utopia, since its value to us depends in no way on any sort of scarcity. Even in Utopia we

37 Suits, *The Grasshopper*, 8.

could voluntarily impose constraints upon ourselves to overcome them just for the sake of overcoming them. The instrumental values of games might no longer be salient, but the noninstrumental value of achievement and of taking on an engaging challenge would remain intact. Hence the distinctive value of games rests not in their providing a value that cannot in principle be found elsewhere, but in their providing the relevant values in an especially *resilient* manner. Indeed, Suits argues that we could reintroduce many of the activities that here and now are primarily instrumental activities *as games*.

This resilience reflects something almost magical about games. With games, we can take some end that otherwise has no intrinsic value—a checkmate position in chess, getting a golf ball into a hole—and transform it into something valuable by incorporating it in the right way in a well-designed game. This ability may tell us something important about the kind of agents we are. We find it easy to enjoy engaging with challenges. This enjoyment can be undermined or tempered if the stakes are too great, of course. Overcoming a challenge when your life is at stake will be stressful rather than rewarding. However, in the context of a game, we are, in the ideal case, anyway, able to lose ourselves in the game and enjoy the challenge it presents. Indeed, we are sometimes even able to enter into “flow states”:

Human development scholar Mihaly Csikszentmihalyi ... has argued that people at play frequently attain a level of experience he calls “flow,” when they become so deeply involved in what they are doing that they lose track of such typical concerns as time of day, external happenings, personal anxieties, and even the sense that they are separate from the activities in which they are acting ... a satisfying experience that results when the player’s skills and commitments mesh almost perfectly with the situations in which they find themselves. Players ... are enticed—and entranced—by appropriate levels of challenge.³⁸

On some accounts, this ability to become absorbed in the challenges is distinctive of our species, and can be given an evolutionary explanation. As we moved from being hunter-gatherers, we had a lot of time on our hands. At the same time, our evolution as hunter-gatherers had primed us for solving puzzles. Kretchmar suggests this intriguing possibility:

Our ancestors during that era had unprecedented amounts of time on their hands.... Increased efficiency produced needs that could not be fully mental by natural forms of play.... Because play is so powerful and

38 Henricks, “The Nature of Play,” 162.

central to good living, we needed to fortify play at some point in our evolutionary history. We did so, in part, by gaming up life in intellectually impressive and unmistakably provocative ways. Games, in this sense, stand as an enduring tribute to the significance of play.³⁹

It is not hard to see echoes of Suits's hypothesis about the resilience of games as a source of value in Kretchmar's account of the origin of games as a cure for boredom in the face of reductions in material scarcity, a point not lost on Kretchmar himself. Other theorists, drawing on the work of Nietzsche, have gone so far as to suggest that agency itself is partly constituted by a drive not simply to achieve ends, but to achieve ends in the face of obstacles of some kind. This is how Paul Katsafanas, for example, understands Nietzsche's view of the "will to power." Katsafanas even explicitly notes the connection to Suits's account of the value of games.⁴⁰

Whatever the explanation of our ability to relish delicious challenges turns out to be, there can be no doubt that it is an important part of our nature, and that adds to the plausibility of Suits's account of the role of games in Utopia and the associated resilience of games as sources of values. There is something fundamentally hopeful about this account of the value of playing games. Here I find it useful to juxtapose Suits's account with one of the classic responses in philosophical theology to the problem of evil. It is sometimes argued that without the adversity we face as human beings that life would inevitably seem boring or pointless. The natural evils with which we must wrestle on this account provide a cure for boredom and provide our lives with rich meaning. Suits's account of the resilience of the value of games in a world with none of the adversity we face in the actual world provides a powerful counterpoise to this line of argument. There is something depressing and pathetic about a species of creatures so dependent on evils to overcome that they could not have a happy or worthwhile life in a world entirely devoid of such adversity. The point is reminiscent of a chilling line from the film *The Matrix*, in which one of the "agents" explained why the first version of the Matrix failed:

Did you know that the first Matrix was designed to be a perfect human world where none suffered—where everyone would be happy? It was a disaster. No one would accept the program. Entire crops were lost. Some believed that we lacked the programming language to describe your per-

39 Kretchmar, "The Normative Heights and Depths of Play," 9.

40 Katsafanas, *Agency and the Foundations of Ethics*, 175n50.

fect world. But I believe that as a species, human beings define their reality through misery and suffering.⁴¹

Suits's account provides a much more positive and hopeful vision of us—one of agents who can create valuable activities to make our lives worthwhile without any need for the horrors of disease and natural disaster. Although the account of playing games and its value developed here differs from Suits's account as well as Hurka's and Tasioulas's accounts, it still coheres well with this profound point in Suits about the resilience of the value of playing games and what it tells us about humanity.

V

In this paper, I have argued that Suits, Hurka, and Tasioulas are right that playing games can have noninstrumental value, and that achievement, taking on an engaging challenge, and play (full stop) all have a role to play in explaining this value. I have further argued that we need to more carefully distinguish the value of achievement from the value of taking on an engaging challenge, as these provide different but related values, each of which appropriately plays a role in different normative contexts. I have argued that both Hurka's view and Tasioulas's view of the explanatory relations between the value of achievement and the value of play (full stop) are oversimple. Neither play nor achievement is without qualification the more fundamental value. Rather, we should endorse what I have called a "variable priority" view of the relation between these values. This more complex view may still leave it a mystery how, if at all, games are an especially distinctive source of value. I have argued that they are distinctive, but not in providing a kind of value not available (or not easily available) in any other context, but instead because they are an especially *resilient* form of value. I concluded by remarking on how to this extent I agree with Suits, and how this provides a more hopeful vision of humanity than one that did not recognize this resilience.⁴²

University of Edinburgh
 mridge@staffmail.ed.ac.uk

41 Wachowski and Wachowski, *The Matrix*.

42 Many thanks to two anonymous referees for their helpful comments on an earlier draft of this paper.

REFERENCES

- Bradford, Gwen. *Achievement*. Oxford: Oxford University Press, 2015.
- . “Evil Achievements and the Principle of Recursion.” In *Oxford Studies in Normative Ethics*, vol. 4, edited by Mark Timmons, 79–97. Oxford: Oxford University Press, 2013.
- Brown, Stuart. *Play: How It Shapes the Brain, Opens the Imagination and Invigorates the Soul*. Penguin: New York, 2010.
- Burke, Richard. “‘Work’ and ‘Play.’” *Ethics* 82, no. 1 (October 1971): 33–47.
- Dancy, Jonathan. *Ethics without Principles*. Oxford: Oxford University Press, 2004.
- Finnis, John. *Natural Law and Natural Rights*. Oxford: Clarendon Press, 1980.
- Griffin, James. *Well-Being: Its Meaning, Measure and Moral Importance*. Oxford: Clarendon Paperbacks, 1986.
- Henricks, Thomas. “The Nature of Play.” *American Journal of Play* 1, no. 2 (Fall 2008): 157–80.
- Huizinga, Johan. *Homo Ludens: A Study of the Play Element in Culture*. 1949. Brooklyn, NY: Angelico Press, 2016.
- Hurka, Thomas. “Games and the Good.” *Aristotelian Society Supplementary Volume* 80 (2006): 217–35.
- . “Two Kinds of Organic Unity.” *Journal of Ethics* 2, no. 4 (1998): 299–320.
- Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press, 2013.
- Korsgaard, Christine M. “Two Distinctions in Goodness.” *Philosophical Review* 92, no. 2 (April 1983): 169–95.
- Kretchmar, R. Scott. “The Normative Heights and Depths of Play.” *Journal of the Philosophy of Sport* 34, no. 1 (2007): 1–12.
- Langton, Rae. “Objective and Unconditioned Value.” *Philosophical Review* 116, no. 2 (April 2007): 157–85.
- McKeever, Sean, and Michael R. Ridge. *Principled Ethics: Generalism as a Regulative Ideal*. Oxford: Oxford University Press, 2006.
- Nussbaum, Martha C. “Human Functioning and Social Justice.” *Political Theory* 20, no. 2 (May 1992): 202–46.
- Ridge, Michael. “How to Play Well with Others.” In *Games, Sport and Play*, edited by Thomas Hurka, 74–98. Oxford: Oxford University Press, 2019.
- Scanlon, T. M. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.
- Smith, Peter K., and Ralph Vollstedt. “On Defining Play: An Empirical Study of

- the Relationship Between Play and Various Play Criteria." *Child Development* 56 (1985): 1042–50.
- Stubbings, David. "Owner of World's Longest Nails Has Them Cut after Growing Them for 66 Years." Guinness World Records, July 11, 2018. <https://guinnessworldrecords.com/news/2018/7/owner-of-worlds-longest-nails-has-them-cut-after-growing-them-for-66-years-532563>.
- Suits, Bernard. *The Grasshopper: Games, Life and Utopia*. Edinburgh: Scottish Academic Press, 1978.
- . "Words on Play." *Journal of the Philosophy of Sport* 4, no. 1 (1977): 117–31.
- Tasioulas, John. "Games and the Good." *Aristotelian Society Supplementary Volume* 80 (2006): 237–64.
- Wachowski, Lana, and Lilly Wachowski, dirs. *The Matrix*. 1999; Burbank, CA: Warner Bros.

OBJECTIFIED WOMEN AND FETISHIZED OBJECTS

Paula Keller

A FAMILIAR CRITIQUE of advertisements from the clothing company American Apparel is that they portray women in objectifying ways. Take this example: in 2012 the brand printed an advertisement of a young woman in a bodysuit, legs apart, looking lasciviously into the camera, with the caption “Now Open,” leaving ambiguous exactly what is open now.¹ This is clearly objectification—but why? The woman is presented as object-like, as inanimate; she is compared to a store. This explanation fits a widespread idea associated with objectification: treating people as if they were objects.

But there is a second sense of objectification: the woman in the American Apparel ad is presented in this object-like way *because* it fits straight-male sexual desires. Those shopping at and those creating American Apparel want to see her naked, open legs; they derive sexual pleasure from those legs. So American Apparel advertisements make her that way in their pictures. This is a second sense of objectification: the content of sexual desire is projected onto women; one then thinks that women *are* the way one sexually desires them to be. What was formerly subjective desire becomes belief about objective reality. We can call this an *objectification* of desires. Projection is an idea familiar from other areas of philosophy: we see colors in the world because we project them onto objects, some have argued in the philosophy of mind.² We discern moral value in the world because we project it onto the world, J.L. Mackie has argued in ethics.³ Taking this idea to feminist philosophy, projection in American Apparel adver-

- 1 The brand has since gone bankrupt and in 2017 reopened under a new image. It now advertises with inclusivity in skin color and body shape: “Introducing our first-ever inclusive nude collection: a celebration of authentic, diverse representation. NUDES encompasses a broad spectrum of skin colors, with sexy premium essentials in 9 shades of nude.” American Apparel, accessed January 10, 2021, <https://americanapparel.com/women/nudes>.
- 2 Most notably, Locke, *An Essay Concerning Human Understanding*, bk. II, ch. 8, secs. 9–10.
- 3 Mackie, *Ethics*, ch. 1.

tisements is about how sexual desires shape how American Apparel customers and creators think about and represent women—here projection is a mechanism from desire to belief.

So we get two related ideas of objectification: treating as objects and forming projective beliefs. Catharine MacKinnon captures both in saying that

like the value of a commodity, women's sexual desirability is fetishized: it is made to appear a quality of the object itself . . . inherent, independent of the social relation that created it.⁴

Here is how this captures the two ideas above. First, MacKinnon speaks of women as objects—the first idea above of treating as objects. I will call this the *moral* idea. Second, MacKinnon specifies that the “social relation” of one sexually desiring women is projected onto the objects, women. This is the second, *epistemic* idea. MacKinnon also adds a third dimension: sexual objectification typically happens to women. It is not a personal but a *political* phenomenon.

A liberal camp around Martha Nussbaum is interested mostly in the moral idea; a radical, Marxist camp around Catharine MacKinnon and Sally Haslanger is concerned with the political idea; and a third group of feminists like Rae Langton and also Haslanger note the epistemic idea.⁵ How do these three senses fit together? For Lina Papadaki they are rivals. For Kathleen Stock they are entirely different projects: Nussbaum's account is said to capture the ordinary, folk usage of the term “objectification.” Epistemic and political accounts are said to describe a problematic phenomenon in the world that, for classificatory purposes, is labeled “objectification.”⁶

I disagree with both Stock's and Papadaki's outline of the field. Instead, I suggest that MacKinnon's quote above can reveal a relation between the three senses: commodity fetishism can function as a model for how to make sense of their relation. This Marxist concept also comprises a moral, political, and epistemic sense and so allows us to read these three senses as pointing out three aspects and three wrongs in one phenomenon—not three different phenomena nor three different claims of what is essential to one phenomenon. My aim is not to defend particularities of each idea of objectification—instead I first collect them to then focus on their connection. My point is not only that all three senses are aspects of one phenomenon, but rather that moral, political, and epistemic fac-

4 MacKinnon, *Toward a Feminist Theory of the State*, 123; see also MacKinnon, *Feminism Unmodified*.

5 See Haslanger, “On Being Objective and Being Objectified,” 241–66.

6 Papadaki, “What Is Objectification?”; Stock, “Sexual Objectification.”

tors interact—without any one of them, we cannot fully explain the widespread existence of the other two. This again is analogous to commodity fetishism.

So what is this talk about commodities and fetishizing? MacKinnon takes it from Marx. Commodity fetishism, in Marx's *Capital*, describes roughly the idea that on the market a product seems to have its value inherently, rather than due to the social relations of production and exchange that really give it value.

Comparing commodity fetishism to sexual objectification is novel because MacKinnon only tentatively suggests this analogy at various places throughout *Feminism Unmodified* and *Toward a Feminist Theory of the State*. We cannot, however, there find a systematic argument for the analogy—MacKinnon leaves work to be done. But what is even more interesting is that nobody has taken up this task: that these two phenomena should be similar is not only relevant for a hybrid account of objectification, but helpful for understanding how self-sustaining social systems work. Sexual objectification, or so I argue, comprises moral, political, and epistemic mechanisms to uphold a patriarchal structure, and the same is true for commodity fetishism and capitalism.

My argument will also show that there is more to sexual objectification than merely treating a person as an object. This is relevant especially since the moral sense—the much more widely accepted sense—focuses only on this definition.⁷ We will see that this moral sense is an oversimplification that leaves questions about the functioning and consequences of sexual objectification unanswered. My project further fits into a wider debate between Marxists and feminists about which of the two frameworks to apply or how to reconcile them. If my argument works, it suggests a general aptness to use Marxist concepts for feminist projects. “Fetishism” might be one example; “alienation” and “ideology” are others.⁸

So what is the analogy between commodity fetishism and sexual objectification? First, sexual objectification reinforces patriarchal *social structures*.⁹ Second, male sexual desire generates a belief about women: the woman in the advertisement is represented as object-like because American Apparel and its customers

7 Although Nussbaum's moral account is already twenty-five years old, it is still a standard in debates on objectification. Here is just a small selection of feminist philosophers working with this account: Papadaki, “Sexual Objectification”; Saul, “On Treating Things as People”; Marino, “The Ethics of Sexual Objectification”; and Langton, *Sexual Solipsism*.

8 For “alienation” see Bartky, *Femininity and Domination*; for “ideology” see Haslanger, “But Mom, Crop-Tops Are Cute!”

9 Following MacKinnon, sexual objectification is what solely defines gender and gender inequality. I am not committed to this strong thesis and will rather follow Haslanger's assumption of a plurality of factors that constructs gender inequality in “On Being Objective and Being Objectified.” For more on this and the indebtedness of my account to Haslanger's see section 3 below.

want her to be. Those groups think that the portrayed woman really has these object-like features *inherently*: they do not think that the origin of this thought is their own sexual desire.¹⁰ Third, the woman in the picture really looks object-like: the picture is *evidence* for the belief that she is object-like. Similarly, commodity fetishism reinforces capitalist *social structures*, involves agents forming the belief that commodities have value *inherently*, and produces *evidence* supporting this belief. So sexual objectification and commodity fetishism are analogous because both involve the following features:

1. Social structure
2. False belief about inherent property
3. Evidence production

That both phenomena reinforce (1) a social structure makes them alike in their function. This paves the ground for a further comparison regarding moral, political, and epistemic aspects: what at first sight looks overly epistemic—speaking of (2) belief and (3) evidence—will turn out to also entail moral and political senses. This analogy allows us to see that an adequate account of sexual objectification must be a hybrid account recognizing moral, political, and epistemic aspects and wrongs involved in the phenomenon.

To show this I first present the moral, political, and epistemic senses of sexual objectification and make some suggestions regarding their possible relationships (in section 1). In section 2, I outline what commodity fetishism is. Sections 3 through 5 explain the analogy with features 1–3 above.

1. THEORIES OF OBJECTIFICATION

1.1. Moral Objectification

The moral sense of sexual objectification best fits public discourse about objectification. Nussbaum gives the canonical account: “In all cases of objectification what is at issue is a question of treating one thing as another: one is treating *as an object* what is really not an object, what is, in fact, a human being.”¹¹ This

10 I follow MacKinnon (quoted above) in her use of “inherently” to mean that a particular property comes with the object. The relevant contrast is that a property is endowed onto the object by a social relation. Consider these examples: wine has the *inherent* property of being a liquid (at room temperature), while—for Catholics—it also has the property of being holy in certain contexts when it signifies Jesus’s blood. This latter property is due to a social relation—it is not inherent.

11 Nussbaum, “Objectification,” 256–57. This basic idea of objectification is derived from Kant, *Lectures on Ethics*.

need not mean that one is treating another as if they were an actual object, say a pen, a painting, or a zucchini. It means rather that aspects of one's behavior or one's attitude toward another person resemble behavior one might also display toward a pen one uses for writing, a painting one marvels at, or a zucchini one buys at the supermarket.

But what makes objectification *sexual* objectification, what exactly does it mean to treat as an "object," and how is this morally dodgy? Nussbaum's initial definition is in need of specification.

First, treating a person as an object might not always be morally objectionable. Here is Nussbaum's example: "If I am lying around with my lover on the bed, and use his stomach as a pillow, there seems to be nothing at all baneful about this, provided that I do so with his consent" and in an environment of mutual equality and respect.¹² Some pleasurable aspects of sexual life, as Nussbaum further notes, might even depend on objectification: gazing at one's partner like one gazes at a statue or using one's partner for one's own sexual pleasure are to some extent enjoyable features of our sexual life. For something to be morally objectionable objectification, Nussbaum suggests that we need to add that objectification is the primary treatment: if I treat my lover *primarily* as a pillow, that is morally objectionable.¹³ In this case, as opposed to the one-off case from before, the lover's general humanity is either actively denied or passively disregarded. Following Papadaki's explication of Nussbaum, it is this denial/disregard that makes objectification morally objectionable.¹⁴ While Nussbaum is not too explicit on what exactly distinguishes morally permissible from impermissible objectification, perhaps the rough picture is clear enough: both kinds exist and pervasiveness and context determine moral permissibility.

Second, there are several senses of treating someone as an "object." Nussbaum lists seven to which Langton adds three further ones; I highlight three of Nussbaum's senses and merely gesture toward the others.

1. *Instrumentality* involves using someone as a tool for one's own ends. Someone masturbating to the American Apparel advertisement "Now Open" uses the picture of the girl as an instrument.

12 Nussbaum, "Objectification," 265.

13 Nussbaum, "Objectification." How consent features here is unclear: Does it make even primary objectification acceptable? Is all nonconsensual objectification morally bad or can there be mild cases of morally permissible objectification not in need of consent? Nussbaum only suggests that both primary treatment and consent combined with equality and respect are relevant for determining objectification's moral status ("Objectification," 265).

14 Papadaki, "What Is Objectification?" 24–25.

The remaining six features (and Langton's three) have to do with a sexual desire that one's partner be a passive object one can do things to.

2. Objects of this kind *lack autonomy*—Nussbaum's second feature. James Bond movies illustrate this well:

James Bond: In the 2015 film *Spectre*, Q introduces Bond to a new car: "Magnificent, isn't she? . . . Beautiful tricks up her sleeve." Cars are female! Across the Bond movies, car scenes are paired with Bond making out with Bond girls. The viewer cannot help but notice that Bond girls and cars are both female and both things Bond plays with. Cars are not autonomous agents, so neither are Bond girls.¹⁵

3. Objects are *inert* and Langton adds that they do not speak and are mere bodies or appearances.¹⁶ This fits the following ad campaign where beauty, not activity, matters:

Helmet Campaign: The German government in March 2019 made a campaign to motivate cyclists to wear helmets. Their slogan "Looks like shit. But saves my life" was paired with images of nearly naked helmet-wearing women (and some men) posing in beds, not active on bikes.¹⁷

This focus on inertness makes clear that the moral wrong involved in objectification is *not* that one pays exclusive attention to material bodies rather than mental persons. Accusations that feminists working on objectification perpetuate a mind–body hierarchy are therefore mistaken.¹⁸ In a world where our materiality as well as our interdependence on one another is often disregarded, explicit at-

15 James Bond objectifies despite being just a fiction: fictions objectify fictional women, cause objectification of nonfictional women, and constitute objectification of nonfictional women (see, e.g., MacKinnon, *Feminism Unmodified*, ch. 14; Langton, *Sexual Solipsism*, ch. 1). A different description of the James Bond example would be that the car is partially upgraded to person status, while Bond girls are partially downgraded so that they meet in the middle. Saul discusses a similar case using the example of vibrators in "On Treating Things as People."

16 Langton, *Sexual Solipsism*, 228–29. One might here object that treating one as a body need not be bad. In fact, deeming it bad disregards our *embodied* lives and wrongly privileges the mental. But Langton can be defended by pointing out that objectification, according to her, reduces us to "body parts" or appearances like "fine paintings and antiques." The relevant contrast is not between the mental and the bodily, but between living and dead/disassembled/two-dimensional.

17 Agence France-Presse, "German Ministry under Fire over 'Sexist' Bike Safety Ad."

18 For example by Cahill, *Overcoming Objectification*. Once this accusation falls, there seems to me to no longer be a need to replace objectification with a different concept, such as Cahill's derivation.

tention to our embodied, not entirely autonomous nature is not objectionable. But this attention is *not* subject to criticism by feminists working on objectification. What characterizes material feminists' emphasis on one's body is that they thereby mean an "active, viable and autonomous body"—as opposed to an enslaved, used, or forced body.¹⁹ This body deserves recognition. It is precisely these characteristics of bodies as alive, autonomous, with subjective feelings and emotions that objectification harms or denies—and that feminists working on the topic object to. Critiques of objectification do not rank minds over bodies; they rank living, autonomous human beings (with both mind and body) over innate objects or nonautonomous instruments. Materiality and embodiment are hotly debated within feminism, and my quick remarks on the matter do not settle these debates. I, however, hope to have shown how one might think of the two parties presented as allies rather than rivals. The idea of the object-like inferior body (be it as a degradation of women or as general conception of the body) is the common enemy of feminists working on objectification and those working on materiality and embodiment.

Third, we can distinguish sexual from other objectification: the former is objectification in a "sexual context" for Nussbaum.²⁰ A master owning a slave or a capitalist employing a worker are examples of nonsexual objectification.²¹ American Apparel, James Bond, and Helmet Campaign are all examples of sexual objectification, as is the following video advertisement:

Miller Lite: Two women get into an argument about why one should drink the beer Miller Lite. The argument gets physical: they first fight each other in a water fountain, then in a mud pit, pulling off each other's clothes—the pinnacle of the scene is them making out. It becomes clear that the scene springs from the fantasy of two men in a bar, imagining the perfect beer commercial.²²

In this ad the women are instruments for the men's sexual pleasure. This is objec-

19 See Grosz, *Volatile Bodies*, 9. It is important to note that the relevant contrast is *not* between autonomous on the one hand and socially dependent on the other. On this picture autonomy would be, as Cahill rightly notes, "not only a fantasy, but a nightmare" for us socially dependent beings; see Cahill, *Overcoming Objectification*, 23. For classic work that emphasizes the importance of this living, relatively autonomous, yet still socially dependent body and critiques a cultural degradation of this body as opposed to the pure, rational mind, see, for example, Young, *On Female Body Experience*; Bordo, *Unbearable Weight*; and Weiss, *Body Images*.

20 Nussbaum, "Objectification," 251.

21 Nussbaum, "Objectification," 262–65.

22 Cahill, *Overcoming Objectification*, 42–47.

tification in a sexual context, not just objectification *simpliciter*. Miller Lite also shows how sexual objectification might reinforce patriarchal structures—in this case by conveying norms of ideal behavior for women. This introduces a second, political sense of objectification, which I turn to next.

Putting together what we have, let us work with the following *moral sense of sexual objectification*:

X constitutes (moral) sexual objectification iff *X* is a treatment in which a person *a*, in a sexual context, treats another person *b* as if *b* were primarily an object (e.g., an instrument, lacking autonomy or inert).

1.2. Political Objectification

Miller Lite reinforces patriarchal structures. This reinforcement is powerful because, following MacKinnon, sexual objectification is itself structural: Miller Lite is just one part of an “elaborate . . . system” with frequent instances.²³ Sexual objectification becomes an important factor in our social world. The elaborate system produces two classes and a hierarchy between them: a class of objectified and of objectifiers. For MacKinnon, objectification not only disproportionately affects one gender—women—it defines what it means to be of different genders. The class of objectified are “women” and the objectifying class are “men”: “the sex difference and the dominance-submission dynamic define each other.”²⁴ “Women are the things and men are the self.”²⁵

This shows that structural objectification, analogous to moral objectification, also involves a degradation to thing-hood. But, following Haslanger, this account lays particular emphasis on women becoming instruments.²⁶ In this respect it resembles particularly closely Nussbaum’s first sense of objectification. While Nussbaum’s account was vague about the specifics of *sexual* objectification—objectification in a sexual context—the political account is more concrete, giving two ways in which objectification is sexual: women become instruments for sexual pleasure specifically, rather than mere instruments in a sexual context, and the thereby affected gender hierarchy is itself eroticized. So women are not merely tools for men’s sexual pleasure; men also derive pleasure from women being the subordinate class.²⁷

23 MacKinnon, *Toward a Feminist Theory of the State*, 141.

24 MacKinnon, *Feminism Unmodified*, 50, and 40–42.

25 MacKinnon, *Feminism Unmodified*, 123.

26 See Haslanger, “On Being Objective and Being Objectified,” 64–65.

27 MacKinnon, *Feminism Unmodified*, 54; Stock also identifies these two factors for *sexual* objectification (“Sexual Objectification”).

MacKinnon's political sense does *not* mean that all men (or only men) objectify women, rather sexual objectification is something "most men adhere to ... nonconsciously ... [because] it is rational for them."²⁸ For MacKinnon, men play the central (but not exclusive) role in sexual objectification. And sexual objectification works: it creates complying, self-objectifying, and inferior female subjects, as Haslanger notes.²⁹

On this picture, sexual objectification is objectionable not only for the moral reason that it denies or disrespects a person's humanity, but because it reinforces the "subjection of women."³⁰ This is true for any sexually objectifying acts, not just for those that are primary or exclusive. This political harm further means that while not all women might directly experience sexual objectification in the form of degrading treatment by men (although it is likely most will), sexual objectification affects them all, in virtue of their membership in the class of women. Sexual objectification makes all women inferior.

Putting this account together, we get the following:

X constitutes (political) sexual objectification iff X is a system of relations between agents such that women socially count as instruments for sexual pleasure while men count as selves, and this gender hierarchy is eroticized.

But moral and political senses leave questions unanswered: Why exactly is sexual objectification of women such a pervasive social phenomenon? If sexual objectification is either morally or politically bad and at least some or most of the objectifiers are responsible moral and political agents, would we not expect them to reduce or attempt to reduce their objectionable behavior? So far, we cannot really make sense of the social reality of sexual objectification: How are the American Apparel example and Helmet Campaign, James Bond, and Miller Lite possible?

1.3. *Epistemic Objectification*

Langton, MacKinnon, and Haslanger give answers to these questions by highlighting an *epistemic sense of sexual objectification*. The epistemic sense involves

28 MacKinnon, *Feminism Unmodified*, 114.

29 Haslanger, "On Being Objective and Being Objectified," 61.

30 MacKinnon, *Feminism Unmodified*, 124. This point is contested: Papadaki ("What Is Objectification?") and Stock ("Sexual Objectification") understand the objectionable nature of MacKinnon-style objectification to primarily have to do with treating someone as a means to an end and therefore in some way violating their humanity. While I do not deny that this might add to objectification's objectionable nature, I maintain that MacKinnon's focus is on badness due to gender hierarchy.

two ideas: projective beliefs about women and the production of evidence that women are as one believes them to be.³¹ What was formerly subjective desire becomes (belief about) objective reality—it is *objectified*. I briefly outline the epistemic sense focusing on the first idea. I spend more time on the second idea in section 5.

Projection in the context of sexual objectification is belief formation about some person due to sexual desire. Here are some familiar examples of such beliefs that a sexual objectifier might form: women are dressed like this so that I can enjoy looking; they like me hitting on them; they sexually desire exactly what I desire; it is impossible for women not to find me attractive, etc. Projection involves, following Langton, the “capacity to generate a belief” that a particular world state is true without making use of external evidence.³² If one projects “they like me hitting on them” onto women, that means one generates this belief without external evidence (e.g., from the observation of women). Projection in the case of sexual objectification means “viewing certain individuals through the lens of one’s [sexual] desire.”³³

Projective belief formation is not ordinary belief formation. Ordinarily, “beliefs aim to fit the world.”³⁴ (Remember Anscombe’s famous shopping cart detective.)³⁵ James Bond’s belief about what sort of organization Spectre is aims to fit (and is created by observation of) the world with Spectre in it. Projective beliefs do the opposite: they are not created by observation with the aim to fit the world. Rather they are generated, in our case, from a particular sexual desire. The Miller Lite example makes this mechanism explicit: we see two women drinking beer, then they start making out. What we at first mistake for a depiction of reality turns out to be a product of male imagination guided by sexual desire. The lines are blurred between sexual fantasy and reality: Do the men in the commercial really know the difference between these two? Are they fully aware that what they imagine women to be like is not what they are actually like? Or is the scenario’s sexual appeal at least partly due to the fact that the imagined scene is not entirely unlike their (and the heterosexual, male target audience’s) perceived reality?

31 Cahill’s competitor-concept “derivatization” also stresses projection: derivatization amounts to de-subjectifying another. One way of taking away subjectivity is to project properties or characteristics onto them regardless of whether they have those or not: “framing or constructing the feminine in terms of the masculine” (Cahill, *Overcoming Objectification*, 50).

32 Langton, *Sexual Solipsism*, 247.

33 Haslanger, “On Being Objective and Being Objectified,” 60.

34 Langton, *Sexual Solipsism*, 244.

35 Anscombe, *Intention*, 56.

The person who forms the belief via projection will be unaware of its origin in sexual desire. James Bond believes that a Bond girl is primarily a beautiful object to such an extent that he is surprised and unprepared when a Bond girl in *Gold-eneye* attacks him—objects do not have autonomy or subjectivity, so they cannot attack. If Bond found out that projection was the way in which he acquired his belief about Bond girls as objects, he would be suspicious of that belief. Keeping his belief would mean risking his life, as the example showed, and it would make him a bad MI6 agent. If he holds this projective belief and if he is (let us assume) a good MI6 agent, he cannot be aware of the projective nature of his belief. Langton asserts accordingly: “projection must make its origins invisible if it is to be belief at all.”³⁶

Aside from projection, the epistemic sense contains a second idea: Bond will treat women according to his projective beliefs about them. If others do so too, women will be *made* objects (at least socially). This comes back full circle to Bond’s beliefs: Bond observes women counting as objects in the social world; he uses that as evidence for his beliefs about women. I say more about evidence in section 5 as it becomes relevant for my comparison to commodity fetishism.

For now, we can combine an *epistemic sense*:

X constitutes (epistemic) sexual objectification iff (1) *X* is a belief-forming mechanism in which an agent *a* forms a belief that some other person *b* has some inherent properties due to *a*’s sexual desires and *a* is unaware of this causal relation, and (2) *b* might be treated according to this belief such that evidence for *a*’s belief is produced.

1.4. Moral, Political, and Epistemic Senses Combined?

How do these three accounts of sexual objectification fit together? In one case sexual objectification is an action, in the other it is a system, in the third it is a particular kind of belief. These seem to be entirely different things. Is it still the case that all three accounts describe the same phenomenon, or do they rather denote completely different aspects of our social lives? My claim is that there is a core phenomenon of sexual objectification in which all three senses show up. A comprehensive account of sexual objectification must therefore be multifaceted. My comparison with commodity fetishism will show how these facets can be combined. But let us start by understanding the differences and similarities among the three accounts.

First, we often act due to our beliefs; so, in a context of political objectification, epistemic objectification might (paired with an intention) *cause* moral

36 Langton, *Sexual Solipsism*, 262.

objectification. If I believe women to be objects, I will treat women as if they were objects. Here are more concrete examples: if one believes that women are there to please men, one (as a man) will treat them as *instruments* for one's pleasure (Nussbaum's feature). If one believes that it does not matter whether one's girlfriend is a scientist or a philosopher as long as she is pretty, one will treat her as *interchangeable* with other pretty women (also one of Nussbaum's seven features). The treatment, which the *moral* sense talks about, can therefore be a consequence of *epistemic* projective belief. Frequent such treatment further serves to build up the eroticized gender hierarchy the *political* sense refers to.

Second, moral and political senses of sexual objectification can provide evidence for the previously unjustified belief formed according to the epistemic sense. If Bond and the like go around treating women as objects, their treatment can be evidence for others that women are object-like; this makes an objectifying belief justified—I come back to this in section 5.

Perhaps the political sense could be understood as a specification of the moral sense: it adds a systemic context to the moral sense and determines who the objectifying agents are—men. But this picture is inaccurate: the political sense not merely adds context, its main focus is on this context. Compared to the moral sense, the political sense therefore shifts the attention to a different phenomenon: first, we are interested in actions, then we are interested in social structure. Surely, the first somehow enacts the second: a society with political objectification must to some extent also feature moral objectification. But the political sense is not simply a subclass of the moral sense.

Looking at the relations between all three senses of objectification, we can see that there are often no tight connections. First, moral objectification can exist without political objectification—take Nussbaum's politically innocent pillow example and (to make it morally objectionable) imagine this were Nussbaum's primary treatment of her lover. Because it goes against the dominant direction of men objectifying women, it does not fit the political sense.

Objectification in the epistemic sense need not be objectification in the moral and political sense. Take this example:

Ann's Perfect Boyfriend: Ann sexually desires boyfriends who are kind and loving. But Joe, her boyfriend, is not like that. Ann desires Joe to be loving and kind to such an extent that she starts believing that Joe is loving and kind and treats him accordingly.³⁷

This is an example of the epistemic sense: Ann projects a property onto Joe that Ann sexually desires Joe to have. But it is also a property that Joe does not have.

37 James describes a similar process in *The Will to Believe*, 23.

Ann is unaware that her belief was formed via projection. Yet this example has nothing to do with the moral sense of sexual objectification: Ann at no point treats (or could from her projective belief about Joe come to treat) Joe as an object in a morally objectionable way, nor does she thereby participate in upholding an elaborate system of the objectification of women. So some cases of projective objectification are morally and politically innocent.

Conversely, some cases of moral and political objectification are not projective—they are epistemically innocent:

Sadistic Rape: This is rape “where non-consent is actively sought, rather than disregarded or ignored. In this sort of case, it’s not that he doesn’t listen to her saying ‘no’—he wants her to say ‘no.’”³⁸ The rapist wants to violate and degrade his victim to the status of an object. He does *not* believe that she already is an object—he makes her one deliberately.

While Sadistic Rape is morally appalling sexual objectification and partakes in a system of women’s objectification, it is not an example of epistemic projective belief formation: the rapist treats his victim as an object but *not* because he desires and therefore believes that she is one. Instead, he knows full well that she is a person and that he violates her; this violation is what he sexually desires.

I showed that moral, political, and epistemic senses relate to one another like overlapping circles in a Venn diagram. But which sense describes what sexual objectification really is? All do! To see this I turn to the comparison with commodity fetishism.

2. COMMODITY FETISHISM AND THREE KINDS OF VALUE

I first ask what exactly it means to speak of a fetish in this case. Then, following Marx, I distinguish three kinds of value. This distinction allows us to understand why, for Marx, commodities are fetishized. Note that it is not my aim to convince you of the existence of commodity fetishism. Even if Marx was wrong about it, the result that sexual objectification works like Marx thought commodity fetishism worked will still be interesting: it will show how a social phenomenon can have moral, political, and epistemic aspects simultaneously. But for now let us assume that commodity fetishism is indeed an existing phenomenon in capitalism. The explanation I give here of commodity fetishism *cannot* do full justice to Marx. My aim is to supply a clear reconstruction that will be useful for testing the analogy with sexual objectification.

First, what is a fetish? A dictionary entry states: a fetish is “an inanimate ob-

38 Langton, *Sexual Solipsism*, 234.

ject worshipped . . . on account of its supposed inherent magical powers, or as being animated by a spirit.”³⁹ For Marx, capitalism turns commodities into fetishes. Commodities are useful objects or services that are bought and sold in market transactions.⁴⁰ In capitalism such commodified objects or services are additionally fetishized: elevated to have inherent “magical” value. But sexual objectification does not elevate anything, one might say—rather it *degrades* women. So where is the analogy?

It has to do with *how* fetishism and *how* sexual objectification operate. Commodity fetishism specifies how objects acquire their fetish character—it specifies a mechanism. This mechanism, I argue, is analogous to sexual objectification. MacKinnon, as quoted in the introduction, talked about “the value of a commodity.” Commodity fetishism is apparently about something that happens to this value. But what kind of value does she mean? Value as socially necessary labor time. To understand this, we need a small detour via three kinds of value that Marx distinguishes: use-value, exchange-value, and value as socially necessary labor-time. Note that *all* of these are substantially different from value in the moral sense.

First, use-value describes the “usefulness of a thing.”⁴¹ Here is an example: a table with legs of the same length will be more useful than a table with legs of different length. It might, however, be that the second table took more time and skill to produce. This has no impact on its use-value: that “is independent from the labor required to appropriate its useful qualities.”⁴²

How do we trade useful commodities for one another? Via exchange-value: this “appears first of all in the quantitative relation, the proportion, in which use-values of one kind exchange for use-values of another kind.”⁴³ I can exchange the tables above with a pearl bracelet.

But how do we determine the ratio of exchange? It cannot be via use-value: following Marx, it is impossible to compare the usefulness of tables with the usefulness of pearl bracelets—they are useful for very different things.⁴⁴ We might then suspect that exchange-value is fixed by the amount of labor that was invested in producing it—Marx calls this “concrete” labor. But that cannot be right either: if I produce a table, this will take at least twenty times longer than if a carpenter does so. But if my table and her table enter the market, my table will

39 Carver, citing the *Oxford English Dictionary*, in “Marx’s Commodity Fetishism,” 51.

40 Cohen, “The Labour Theory of Value and the Concept of Exploitation,” 339.

41 Marx, *Capital*, 1:126.

42 Marx, *Capital*, 1:126.

43 Marx, *Capital*, 1:126.

44 Marx, *Capital*, 1:128.

not suddenly be twenty times more expensive than hers. Exchange-value cannot reflect actual labor-time.

What is it then? Marx introduces value—the value MacKinnon speaks of in the quoted passage in the introduction. Value is determined by “socially necessary labour-time” and exchange-value is a representation of value.⁴⁵ Socially necessary labor-time “is the labour-time required to produce any use-value under the conditions of production normal for a given society and with the average degree of skill and intensity of labour.”⁴⁶ This means that value does not orient itself on concrete labor that went into a product, but on an abstraction. “The amount of time *actually* spent producing it [is] . . . strictly irrelevant to its value.”⁴⁷

But this also means, following Marx, that we come to think of value as *inherent* in the commodity itself: we observe that value cannot have come from concrete labor, so it must come from elsewhere. To us the only viable option, following Marx, seems to be that it comes from the commodity itself.⁴⁸ This belief about inherent value in a commodity is false: a commodity does not have inherent value somehow within it, “so far no chemist has ever discovered” that.⁴⁹ Rather the social relation of production—socially necessary labor-time—determines value. Looking back to MacKinnon’s quote we now understand why she writes that “the value of a commodity . . . is made to appear a quality of the object itself, inherent, independent of the social relation that created it.”⁵⁰

We might at this point wonder why commodity fetishism is so interesting to Marx: sure, it involves me holding a false belief—but only about the origin of value, not about the amount of the property’s value itself. After all, the commodity really has value. I am just mistaken about where exactly this value came from: from social relations, not out of the commodity itself. But this mistake in character is bad in at least three ways: it prevents us from knowing something we ought to know (the epistemic way), it enslaves producers (the political way), and it makes us treat others as objects (the moral way). As we will see, sexual objectification shares exactly these three problems with commodity fetishism.

45 Marx, *Capital*, 1:129.

46 Marx, *Capital*, 1:129. This bit of Marx, the labor theory of value, is controversial: value might also be affected by things other than socially necessary labor time, such as the existence of a monopoly driving up exchange value, as noted by Cohen, “The Labour Theory,” 350–51.

47 Cohen, “The Labour Theory,” 345, emphasis added.

48 Marx, *Capital*, 1:164.

49 Marx, *Capital*, 1:177. One difficulty for this conception is presented by commodities into which no labor went, e.g., virgin soil. Their value is determined by the social relations of exchange, but cannot be cashed out in terms of socially necessary labor time.

50 MacKinnon, *Toward a Feminist Theory of the State*, 123.

But to draw conclusions for sexual objectification, we first need to make sure we are drawing our conclusions from the right comparison. Why commodity fetishism, why not, say, alienation? Alienation might also have moral, political, and epistemic aspects: alienated from others, we might cease to employ empathy and morally wrong our co-citizens; a society of alienated individuals might be an undesirable political order; and being alienated from others involves making the epistemic error that others are unlike me or that I have no influence on them. So alienation too has moral, political, and epistemic aspects; why not look to alienation for a model for sexual objectification?

For two reasons: first, alienation's moral, political, and epistemic aspects have no clear social function of upholding a social structure as is the case for sexual objectification. Instead, for alienation, the three aspects seem to be separate consequences of one phenomenon. When it is already unclear how these three aspects fit together there, looking at alienation will not help to answer this question of fit for sexual objectification. Second, alienation's moral, political, and epistemic aspects are very different from sexual objectification's parallel aspects: a moral disregard for humanity, a politically unjust hierarchy, and an epistemically flawed belief in inherent properties. Alienation, it seems, would not help us in constructing a combined account of sexual objectification.

Both reasons do not hold for commodity fetishism. First, commodity fetishism and sexual objectification are functionally alike: they are both playing a part in maintaining the status quo. I show this analogy in social structure in section 3. Second, commodity fetishism also contains a false belief about inherent property (section 4), a moral disregard for humanity, and a politically unjust hierarchy (section 5). This means that sexual objectification and commodity fetishism are also alike regarding the content of moral, political, and epistemic aspects.⁵¹

3. ANALOGY IN SOCIAL STRUCTURE

Commodity fetishism and sexual objectification both occur within larger social structures of capitalism and patriarchy and have a status quo-maintaining role to play in them. This similarity in function means that, when comparing the two phenomena, we are comparing like with like.

Above I noted that sexual objectification (following MacKinnon) and com-

51 Further analogies might be drawn: between women's beauty and money as fetishized universal commodities, or between workers as creators of valuable commodities and women as sexual agents—the parts workers and women play in commodity production and sexual life is obscured. As these additional analogies are not essential to my main project of understanding the relation between three accounts of objectification, I will not pursue them further.

modity fetishism are themselves structural: they are systems constituted by a multitude of single actions or beliefs, not merely *a single* action or belief. Here, I am interested in how both phenomena relate to larger social structures: capitalism and patriarchy. Sally Haslanger defines a social structure as “a network of social relations, some of which are to other people, some of which are to non-human animals, some to things.”⁵² A social relation, as Haslanger explicates in her writings on objectification, specifies an “extrinsic property of individuals” and “depends upon the organisation of social life.”⁵³ A social relation contrasts with natural relations or properties: to borrow one of Haslanger’s examples, being a scapegoat is clearly social; it has nothing to do with the scapegoat’s natural properties. We can add that for a network of such social relations to be a social structure it must exist over at least some time and within some social group. It then becomes a social structure *of* this particular social group. Exactly how long such relations must persist, how large the social group must at least or at most be, and how (in)frequently the social relations must be reaffirmed are questions for further debate. For my project a preliminary definition suffices. I will therefore say that a social structure is a persisting network of social relations within a social group.

Commodity fetishism supports class relations in capitalism; sexual objectification (as MacKinnon’s political sense stresses most explicitly) reinforces patriarchal relations.⁵⁴ Capitalism contains a network of social relations between a class of owners of the means of production and a class of workers laboring for this first class.⁵⁵ These relations are typically relations of power inequality and domination. Commodity fetishism means, as I demonstrated, that the worth of a commodity is seen to be inherent in it, while it is actually produced by relations of production and exchange. This means that large parts of economic inequality will appear as arising out of the inherent property of a commodity. They will seem to be unrelated to human social relations and out of reach for humans to change: the commodity that is one’s specific labor is inherently worth some limited amount, but not enough to afford another commodity, such as a house in a well-thought-of district. In this case commodity fetishism makes it so that one’s social situation seems unalterable—the commodities that enable or prevent opportunities have a fixed inherent value. So commodity fetishism supports the pertaining capitalist social relations.

52 Haslanger, “What Is a (Social) Structural Explanation?” 128.

53 Haslanger, “On Being Objective and Being Objectified,” 41.

54 This relation to patriarchy is also assumed in Nussbaum, “Objectification,” 249–50, 271; and Langton, *Sexual Solipsism*, esp. 241–45.

55 Cohen, *Karl Marx’s Theory of History*, chs. 3, 5.

The same is true for sexual objectification within patriarchy. Patriarchy is a system of social relations such that “men constitute the dominant social group” and women the subordinate social group.⁵⁶ Sexual objectification is one such relation: according to the moral and political senses specified in sections 1.1 and 1.2, men relate to women by treating them primarily as objects. According to the epistemic sense from section 1.3, this has to do with men’s sexual desires. This degrading treatment reinforces the existence of men as the dominant social group, women as the subordinate social group, and according to the political sense, even eroticizes this order. In short, objectification reinforces patriarchy.

So both commodity fetishism and sexual objectification reinforce capitalist and patriarchal social structures. This gives us a starting point for a further analogy: the two phenomena work similarly in our social world. Both involve agents forming false beliefs about inherent properties of some x (section 4). This mechanism is so successful because it generates additional evidence for the formed beliefs (section 5). So the two phenomena not only are alike in overall social function but also in how they achieve this function.

In this section I already employed all three senses of sexual objectification to make clear its relation to patriarchy. But how is it that each aspect of objectification exists in our social world? I argue that they form a network to mutually enforce one other—this is also true of commodity fetishism and its moral, political, and epistemic aspects.

4. FALSE BELIEFS ABOUT INHERENT PROPERTIES

4.1. *Commodity Fetishism and False Beliefs*

To see how, in both cases, false beliefs about the inherent properties of some object or person are formed, I first turn to commodity fetishism, then to sexual objectification. In commodity fetishism, one forms the belief that commodities have their value inherently when it is really endowed onto them by the social relations of production and exchange.

How exactly is this false belief generated? In a supermarket “it is impossible to know anything about the labor or the laborers [who congealed value in the lettuce]... You cannot, for example, figure out in the supermarket whether the lettuce has been produced by happy laborers, miserable laborers, slave laborers, wage laborers or some self-employed peasant.”⁵⁷ This is so because capitalism features private production and public exchange, following Marx: commodities

⁵⁶ Chambers, “Feminism,” 562.

⁵⁷ Harvey, *A Companion to Marx’s Capital*, 39–40.

are “the products of the labour of private individuals who work independently of each other. . . . The producers do not come into social contact until they exchange the products of their labour.”⁵⁸ This means that producers only interact with one another to exchange products once these are finished commodities. As a carpenter I only approach a pearl-bracelet maker to exchange some tables for a bracelet once my tables exist as finished products. But *from* my finished table neither I nor the pearl-bracelet producer nor some other agent can tell whether I worked on it for eight days or eight months. Sure, I know how long I worked on that table—but I know that from doing the work, not from looking at the finished product. This means that capitalist private production and public exchange make information about actual labor time epistemically inaccessible on the market.

But what is the relation between ignorance about actual labor time and the fetishizing belief that value is inherent in commodities? Since we cannot exchange commodities by comparing actual labor time involved in production and exchange, something else—we infer—must determine their exchange-value. But as actual labor time—coming from outside to the commodity—is not available, the determining factor must instead—it seems to us—come from the commodity’s inside. So we come to believe, as I showed in section 2, that commodities have inherent value that determines exchange-value.

Summing up, agents in capitalism form beliefs about the inherent value of a commodity because the finished product (e.g., in the supermarket) does not speak of the social relations that endow value onto the commodity. It is this epistemic mechanism to which we will find an analogous mechanism in the case of sexual objectification. Objectification’s epistemic sense outlined above, to do with projection, will have a role to play here.

4.2. *Sexual Objectification and False Beliefs*

What is the false belief in the context of sexual objectification? Sexual desire produces a projective belief about the inherent object-like character of women, following in particular the epistemic sense of sexual objectification. Let us look in more detail at how this false-belief formation works—it will work *differently* than in commodity fetishism, but will produce the *same* result: a false belief about an inherent property.

Projective belief is produced from sexual desire. I take it that sexual desire can be described as a desire for sexual pleasure. And desires for x are generally desires for x to obtain—philosophers of mind therefore speak of desires having

propositional content.⁵⁹ So my desire for sexual pleasure is my desire that I be in a state of sexual pleasure. We can further distinguish intrinsic and instrumental desire. Sexual desires that women are beautiful objects or conform to other sexual fantasies might be instrumental desires: one desires their satisfaction instrumentally so that one experiences sexual pleasure.⁶⁰ It is these instrumental sexual desires that are the subject of projection.

Desires typically dispose one to act so that x obtains, following what philosophers of mind call action-based theories of desire: I have a desire for tea, so I am disposed to go and make myself a cup of tea.⁶¹ Some sexual desires are action-based desires: I desire to have sex with my partner, which makes me disposed to ask them whether they also want to have sex. But desires in sexual objectification are different. They are projective, generating a belief without any external evidence for this belief. Here, “belief [rather than action] is driven by desire.”⁶²

As an aside, we might wonder about the status of these desires: Is one born with them, or does one socially acquire them? The desire for tea is partly social: while we all need to drink, a taste for tea partly depends on what beverages one is used to in one’s culture. In the case of sexual desire, we can ask the same question: Are men’s sexual desires to objectify women natural, social, or a mix of both? I follow MacKinnon in holding that they are at least to a large extent social.⁶³ In a non-patriarchal society, what we want sexually might be very different.

So both commodity fetishism and sexual objectification involve agents forming false beliefs about the inherent properties of commodities and women—this is the epistemic component they share and the reason why both phenomena are epistemically bad. In an alternative, socialist world, an epistemic ill of commodity fetishism would not exist: there, “an association of free men, working with the means of production held in common ... [has] *full self-awareness* as one single social labour force.”⁶⁴ In this socialist utopia all free people would know about the social character of commodity value, while in capitalism they do not.

This epistemic ill is mirrored in sexual objectification: women are falsely believed to *be* object-like. We can point to a further epistemic ill in this case: pro-

59 Schroeder, “Desire.”

60 Whether the agents who hold such instrumental desires will be aware of their instrumental nature is a different matter. It might well be that the objectifying, originally instrumental desire has acquired the phenomenology of an intrinsic desire.

61 Schroeder, “Desire.”

62 Langton, *Sexual Solipsism*, 244.

63 MacKinnon, *Toward a Feminist Theory of the State*, 131–32.

64 Marx, *Capital*, 1:171, emphasis added.

jective belief violates the rules for direction of fit. “Belief is supposed to fit the world.”⁶⁵ But in sexual objectification, as I have described it, belief fits sexual desire. Commodity fetishism lacks this second epistemic ill—does this harm the analogy?

4.3. Marx’s Comparison: Fetishism in Religion

Objectifying beliefs, we might worry, seem very different from fetishizing beliefs: projection in one case has nothing to do with inference from the organization of production and exchange in the other. This difference in belief formation makes a possible comparison between sexual objectification and commodity fetishism thin and uninteresting, one might claim.

But Marx compares commodity fetishism to religion: religion has its own fetishism.⁶⁶ And religion contains projective beliefs just like sexual objectification—so the analogy with sexual objectification is not so far-fetched, at least not for Marx. He writes: in “the misty realm of religion . . . the products of the human brain [i.e., gods] appear as autonomous figures endowed with a life of their own, which enter into relations both with each other and with the human race.”⁶⁷ In religion, following Marx, we create gods, they are “products of the human brain.” This means they are generated without making use of external evidence. And these gods appear to us as “autonomous figures,” precisely not as created by our own brains. We are unaware of their real origin in our own brain. So gods, following Marx, are *projected*.⁶⁸ This means that religious fetishism contains projection just like sexual objectification does. But if religion and sexual objectification share projection, sexual objectification and religion lead one to form beliefs in closely analogous ways. And if religion, for Marx, is analogous to commodity fetishism, then sexual objectification can be too.

Again, I have not done enough to argue for the parallel between religion and commodity fetishism; this would go beyond the scope of this paper—I merely explained why Marx holds that they are parallel. If commodity fetishism and religion are analogous, as Marx thinks they are, then Marx’s commodity fetish-

65 Langton, *Sexual Solipsism*, 266.

66 It would be more historically accurate to say that Marx applies the idea of religious fetishes (already in Hegel, *Vorlesungen über die Philosophie der Weltgeschichte*, 222) to commodities.

67 Marx, *Capital*, 1:165.

68 The story is surely more complicated: religion is fundamentally social and as such not every member of a religious group will create their gods using their own brain—most will take over the gods others have created. This further raises the question of whether the creators might not have created religion intentionally, e.g., in order to justify their own power, rather than unintentionally via projection. Marx’s short discussion in *Capital* is ambiguous on this point (1:165).

ism and sexual objectification can be analogous too. Even though they differ in how exactly they produce false beliefs in inherent properties, this is still their epistemic commonality.

But how exactly can these beliefs continue to exist? Following MacKinnon, sexual objectification works because “men ... [have the] power to force the world to be any way their mind can invent.”⁶⁹ In section 5 I argue that men force the world so that it entails evidence for their objectifying beliefs—and in the case of commodity fetishism the world also comes to entail evidence for beliefs about inherent value. Here moral and political senses come into play and support the epistemic sense.

5. EVIDENCE PRODUCTION

5.1. *Commodity Fetishism and Evidence Production*

Evidence for the belief that commodities have inherent value and, correspondingly, for the belief that women are object-like, reinforces these beliefs. To explain how this evidence is formed I first turn to commodity fetishism, then to sexual objectification.⁷⁰

What evidence is there for the belief that a commodity has inherent value? In a capitalist social world we observe others attaching inherent value to commodities. Here is a concrete example:

Real Estate Agent: Joan is an agent selling land. To determine the price of a piece of farmland she looks at the quality of the soil or the size of the property. This means she treats the land as if value came from itself: as if “ground rent grows out of the soil.”⁷¹ She does not engage in minute research about the various possible human uses of this piece of land.

If I visit Joan during her work, I will find that she treats the land as inherently valuable: value is somehow in the “soil” itself. This treatment of commodities *constitutes* them as fetishes: they gain the social meaning of objects with inherent value once individuals treat them as such. But my observation of Real Estate Agent is evidence for my belief that value is inherent in commodities. Marx has his own example to demonstrate the same point: money. Money, following

69 MacKinnon, *Toward a Feminist Theory of the State*, 122.

70 It is worth noting that the evidence I will talk about might not only reinforce objectifying and fetishizing beliefs—it might also in some cases create them. The story I am telling here necessarily simplifies the existing social reality. The hope, however, is that it does so in illuminating ways.

71 Marx, *Capital*, 1:176.

Marx, becomes the universal measure for value. Exchanging commodities for money is evidence that those commodities have inherent value, quantifiable in a particular currency.⁷²

A political sense of commodity fetishism crystallizes itself: a network of social relations between consumers and commodities forms an elaborate system. This political aspect loops back to the epistemic aspect: evidence for the belief that commodities have inherent value is generated on the market where commodities are traded as if they had inherent value. Real Estate Agent showed what this manner might look like and how this evidence can be gathered.

But the political sense goes further, according to Marx: in commodity fetishism “production has mastery over man.”⁷³ Human creations “come to enslave and oppress their creators.”⁷⁴ This—briefly sketched—is due to the false belief about inherent commodity value. This belief turns commodities into special things such that producers lead lives in which commodities play a central role: “they will [for example] be compelled . . . to perceive given objects solely as ‘things’ that one can potentially make a profit on.”⁷⁵ This hunt for profit enslaves men. Commodity fetishism therefore is not only an elaborate system, analogous to MacKinnon’s analysis of sexual objectification. It is also politically objectionable because it oppresses social groups—again analogous to MacKinnon’s analysis.

Before I trace the moral aspect of commodity fetishism, let us first turn to evidence production for sexual objectification.

5.2. *Sexual Objectification and Evidence Production*

Epistemic sexual objectification produces evidence for its projective beliefs. This evidence can be found in instances of moral and political objectification. Men do not merely believe that women are things—“women are the things.”⁷⁶ What

72 Marx, *Capital*, 1:168–69.

73 Marx, *Capital*, 1:175.

74 Leopold, “Alienation.”

75 Honneth, *Reification*, 22.

76 MacKinnon, *Toward a Feminist Theory of the State*, 120. But if women are things, why is treating them as such still bad? (Papadaki, “What Is Objectification?” 21, mentions this worry.) First, it is still politically bad: women should not be things. And it might be epistemically bad when it involves projection from sexual desires. Second, women are things in the sense that they socially count as things—think back to Haslanger’s scapegoat. But this does not make them things in all senses: women remain human beings with certain (perhaps oppressed, but nevertheless existent) needs and abilities (e.g., nutrition, breathing, movement, action). Women do not become full objects, lacking *any* subjectivity. So objectification will remain morally wrong, even if to a lesser degree after successful previous objectification.

does that mean? Women are things because they are treated as things—Nussbaum's favored moral sense captures this well. This treatment both *constitutes* and *causes* their existence as objects. I take this distinction between constitutive and causal social construction from Haslanger.⁷⁷ For something to be constitutively constructed means that “in defining it we must make reference to social factors”: marriage constitutes one as a wife. Something is causally constructed when “social factors play a causal role in bringing it into existence”: that this wife speaks English is caused by how she learned the language.⁷⁸ Applying this distinction to sexual objectification and commodity fetishism will help us see the relation between epistemic, moral, and political senses.

First, how are women treated as objects? In section 4 we saw that men hold projective beliefs about women—the epistemic sense. But projective beliefs can guide action. James Bond's belief that women are objects for his sexual pleasure might—when confronted with a woman—cause him to act accordingly and treat her as an object—the moral sense. But why does this *constitute* women's existence as objects? If Bond treats his Bond girls like he treats his cars (as objects), this means that the Bond girls come to socially count as car-like objects. Andrea Dworkin seems to have this constitutive relation in mind when she asserts that “those who can be used as if they are not fully human are no longer fully human in social terms; their humanity is hurt by being diminished.”⁷⁹ Usage as inferior, for Dworkin, is the very thing that means inferiority. This is analogous to Real Estate Agent in commodity fetishism: others treating women as objects constitutes them as objects just like others treating value as inherent in commodities constitutes inherent value. This so constituted social reality provides evidence for objectifying beliefs and so the moral sense of sexual objectification loops back to the epistemic sense.

For sexual objectification there is a second sense in which women are objects. (Moral) treatment of women as objects *causes* their existence as objects: as MacKinnon notes, “women have little choice but to become persons who then freely choose women's roles.”⁸⁰ That women are objects becomes “empirically real.”⁸¹ Constant treatment as an object might make me think that this treatment must be right about me: I must be object-like. But once I think that, I will start behaving accordingly. This means that men's objectifying treatment causes women to behave in object-like ways, e.g., paying immense attention to

77 Haslanger, “Ontology and Social Construction,” 98, 103.

78 Haslanger, “Ontology and Social Construction.”

79 Dworkin, “Against the Male Flood,” 31.

80 MacKinnon, *Toward a Feminist Theory of the State*, 124.

81 MacKinnon, *Toward a Feminist Theory of the State*, 230.

their outer appearance or presenting themselves as generally passive, quiet, or shy. Here we start to see how an elaborate system of sexual objectification can set up a gender hierarchy: (epistemic) projective beliefs lead to a particular (moral) treatment that (politically) causes women to become inferior. With this relation in mind, let us look back to commodity fetishism one last time.

5.3. *Three Aspects of Two Phenomena*

Is there *causal social construction* in commodity fetishism? Commodities (as objects) do not think or act at all—so for those this analogy falls down.⁸² But services are also commodities. Here the analogy can work: workers in capitalism might (due to the fetishization of the services they provide) center their lives around their work—this means that they alter their behavior and mindset. This centrality of work in our lives forms part of the political sense of commodity fetishism: the elaborate, enslaving system is not only produced by our relation to nonhuman commodities but also by relating to our services and our labor as a commodity.

Focusing on services as commodities, we can also discern a moral aspect of commodity fetishism, analogous to Nussbaum's account of sexual objectification: individuals are treated primarily as objects.⁸³ How so? One consequence of commodity fetishism is that social relations between producers turn into relations between things.⁸⁴ The exchange of my labor with someone else's labor (so that I acquire a pearl bracelet and they acquire tables) seems to be an exchange between bracelet and tables—not between human labor. This is so because commodity fetishism, as we saw, entails that value appears inherent in a commodity. But in this way commodities gain social significance—to such an extent that individuals think of themselves and others in the same way in which they think of commodities. Axel Honneth observes: “As soon as social agents begin to relate to each other primarily via the exchange of equivalent commodities, they will be compelled . . . to regard each other solely as ‘object’ of a profitable transaction.”⁸⁵ Commodity fetishism brings with it that we treat others as objects.

We have therefore traced political, moral, and epistemic aspects: commodity

82 Marx imagines what they would say if they could speak: “What does belong to us as objects . . . is our value. Our own intercourse as commodities proves it” (*Capital*, 1:176). This is just as false as when women (due to causal social construction) said: “We are object-like. Our own behavior proves it.”

83 Marx himself might have resisted the claim that such treatment is *morally* bad; on Marx and morality see, e.g., Wolff, “Karl Marx.”

84 Marx, *Capital*, 1:165.

85 Honneth, *Reification*, 22.

fetishism draws up an elaborate system of relations to commodities that enslaves especially workers; it makes us disregard others' humanity, viewing them only as exchangeable commodities or as mere means to acquire commodities; and it leads us to hold false beliefs about the inherent properties of commodities.

Note that causal (and to some extent constitutive) social construction for both sexual objectification and commodity fetishism only work if the treatment is frequent, or comes from a large or powerful group. If one little boy occasionally treated me as an object, I would laugh at him. This is different if the boy's treatment mirrors treatment by a large or powerful group. Men in patriarchy are such a group: "men have power."⁸⁶ So MacKinnon was right: men's power has a role to play in sexual objectification—it produces a social reality that corresponds to objectifying beliefs. Here we see how epistemic and moral aspects bring about political aspects of sexual objectification. This is mirrored by commodity fetishism: if only one little boy occasionally fetishized his toy car, commodities would not become inherently valuable. It is different if the little boy's behavior was typical and mirrored by the behavior of real estate agents like Joan, adult car owners, and ordinary consumers. Only then does commodity fetishism become an elaborate and oppressive system, analogous to MacKinnon's political sense.

We began with MacKinnon's analogy that

like the value of a commodity, women's sexual desirability is fetishized: it is made to appear a quality of the object itself . . . inherent, independent of the social relation that created it.⁸⁷

Now we are in a position to fully understand this passage: we know what MacKinnon and Marx mean by value (section 2); we know why commodities are seen to have value attached to them inherently (section 2); we know that commodities' value is really due to social relations (section 2); we also know how object-like women are created by a social relation of men desiring women (section 5); and we know how this social relation is falsely portrayed as the inherent character of women (section 4). And section 3 showed that both phenomena play a status quo-supporting role.

Drawing together the important strings of our analogy, we can see how moral, political, and epistemic aspects of one phenomenon can mutually reinforce the other aspects. The epistemic aspect supported the moral aspect: agents act on their false beliefs about inherent properties and treat women and other workers as object- or commodity-like. The epistemic aspect also contributed to the political aspect: when women are believed to be inferior objects, they (consti-

86 MacKinnon, *Feminism Unmodified*, 119.

87 MacKinnon, *Toward a Feminist Theory of the State*, 123.

tively and causally) become inferior objects; and workers become enslaved by the fixation on commodities and services when belief in their inherent value is widespread. Moral and political aspects relate in that the first contributes to but also instantiates the second: individual cases of objectification are part of an elaborate political system; and individual transactions between agents on the market, where each agent is reduced to an instrument for acquiring commodities, are part of a system that downgrades workers and fetishizes commodities. Both moral and political aspects further function as evidence for the false beliefs the epistemic sense puts center stage: observing American Apparel, James Bond, Helmet Campaign, and Miller Lite, I might believe that women are inherently object-like; just like Real Estate Agent, a visit to the supermarket and using money might convince me of commodities' inherent value.

We can conclude that, when debating over which of the three outlined accounts of sexual objectification to favor, we must understand that all three hang together. They are neither rivals (as Papadaki had it) nor different projects (as Stock saw it). The various interrelating aspects of commodity fetishism provided a model for how to best think about this complicated picture. They also helped rectify the imbalance toward the moral account that debates on sexual objectification seem to have: I showed that the moral account on its own is incomplete as an account of sexual objectification. It cannot explain why objectification occurs and why it occurs so frequently. Paying attention to an interrelated system with moral, political, and epistemic components fares better.

We might further suspect that two analogous phenomena can also be overcome in analogous ways. Both depend on ignorance—about projection's role in objectification and about the market's role in making a commodity appear inherently valuable. Eradicate the ignorance and you eradicate objectification and fetishization. Surely this is easier said than done, but perhaps saying is a first step toward doing.

University of Cambridge
pmk41@cam.ac.uk

REFERENCES

- Agence France-Presse. "German Ministry under Fire over 'Sexist' Bike Safety Ad." *The Guardian*, March 24, 2019. <https://www.theguardian.com/media/2019/mar/24/german-bike-safety-ad-featuring-model-in-bra-and-helmet-sexist>.

- Anscombe, G. E. M. *Intention*. Oxford: Basil Blackwell, 1976.
- Bartky, Sandra. *Femininity and Domination: Studies in the Phenomenology of Oppression*. New York: Routledge, 1990.
- Bordo, Susan. *Unbearable Weight: Feminism, Western Culture, and the Body*. Oakland: University of California Press, 2004.
- Cahill, Ann J. *Overcoming Objectification: A Carnal Ethics*. New York: Routledge, 2011.
- Carver, Terrell. "Marx's Commodity Fetishism." *Inquiry* 18, no. 1 (1975): 39–63.
- Chambers, Clare. "Feminism." In *The Oxford Handbook of Political Ideologies*, edited by Michael Freeden and Marc Stears, 562–82. Oxford: Oxford University Press, 2013.
- Cohen, G. A. *Karl Marx's Theory of History: A Defense*. Oxford: Clarendon Press, 1978.
- . "The Labor Theory of Value and the Concept of Exploitation." *Philosophy and Public Affairs* 8, no. 4 (Summer 1979): 338–60.
- Dworkin, Andrea. "Against the Male Flood: Censorship, Pornography, and Equality." In *Feminism and Pornography*, edited by Drucilla Cornell, 19–38. Oxford: Oxford University Press, 2000.
- Grosz, Elizabeth. *Volatile Bodies: Toward a Corporeal Feminism*. Bloomington, IN: Indiana University Press, 1994.
- Harvey, David. *A Companion to Marx's Capital: The Complete Edition*. London: Verso, 2010.
- Haslanger, Sally. "'But Mom, Crop-Tops Are Cute!' Social Knowledge, Social Structure and Ideology Critique." *Philosophical Issues* 17 (2007): 70–91.
- . "On Being Objective and Being Objectified." In *Resisting Reality: Social Construction and Social Critique*, 35–82. Oxford: Oxford University Press, 2012.
- . "Ontology and Social Construction." *Philosophical Topics* 23, no. 2 (Fall 1995): 95–125.
- . "What Is a (Social) Structural Explanation?" *Philosophical Studies* 173, no. 1 (January 2016): 113–30.
- Hegel, Georg W.F. *Vorlesungen über die Philosophie der Weltgeschichte* [Lectures on the Philosophy of History]. Hamburg: Felix Meiner Verlag, 1955.
- Honneth, Axel. *Reification: A New Look at an Old Idea*. Oxford: Oxford University Press, 2008.
- James, William. *The Will to Believe*. New York: Longmans, Green and Co., 1897.
- Kant, Immanuel. *Lectures on Ethics*. Translated by Louis Infield. New York: Harper and Row, 1963.

- Langton, Rae Helen. *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. Oxford: Oxford University Press, 2009.
- Leopold, David. "Alienation." *Stanford Encyclopedia of Philosophy* (Fall 2018). <https://plato.stanford.edu/archives/fall2018/entries/alienation>.
- Locke, John. *An Essay Concerning Human Understanding*. Vol. 1 of *The Works of John Locke*, 12th ed. London: Rivington, 1824. <https://oll.libertyfund.org/title/locke-the-works-vol-1-an-essay-concerning-human-understanding-part-1>.
- Mackie, J.L. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin, 1990.
- MacKinnon, Catharine A. *Feminism Unmodified: Discourses on Life and Law*. Cambridge, MA: Harvard University Press, 1987.
- . *Toward a Feminist Theory of the State*. Cambridge, MA: Harvard University Press, 1989.
- Marino, Patricia. "The Ethics of Sexual Objectification: Autonomy and Consent." *Inquiry* 51, no. 4 (2008): 345–64.
- Marx, Karl. *Capital*. 2 vols. Translated by Ben Fowkes. London: Penguin Classics, 1990.
- Nussbaum, Martha. "Objectification." *Philosophy and Public Affairs* 24, no. 4 (Autumn 1995): 249–91.
- Papadaki, Lina. "Sexual Objectification: From Kant to Contemporary Feminism." *Contemporary Political Theory* 6 (2007): 330–48.
- . "What Is Objectification?" *Journal of Moral Philosophy* 7, no. 1 (January 2010): 16–36.
- Saul, Jennifer M. "On Treating Things as People: Objectification, Pornography and the History of the Vibrator." *Hypatia* 21, no. 1 (Spring 2006): 45–61.
- Schroeder, Tim. "Desire." *Stanford Encyclopedia of Philosophy* (Summer 2017). <https://plato.stanford.edu/archives/sum2017/entries/desire>.
- Stock, Kathleen. "Sexual Objectification." *Analysis* 75, no. 2 (April 2015): 191–95.
- . "Sexual Objectification, Objectifying Images, and 'Mind-Insensitive Seeing-As.'" In *Evaluative Perception*, edited by Anna Bergqvist and Robert Cowan, 295–310. Oxford: Oxford University Press, 2018.
- Weiss, Gail. *Body Images: Embodiment as Intercorporeality*. London: Routledge, 1998.
- Wolff, Jonathan. "Karl Marx." *Stanford Encyclopedia of Philosophy* (Winter 2017). <https://plato.stanford.edu/archives/win2017/entries/marx>.
- Young, Iris Marion. *On Female Body Experience: "Throwing Like a Girl" and Other Essays*. Oxford: Oxford University Press, 2005.

BLAMING FOR UNREASONABLENESS ACCOUNTABILITY WITHOUT ILL WILL

Alisabeth Ayars

MANY THEORISTS of moral responsibility endorse the principle that X is blameworthy for a wrong act A only if in doing A , X expressed ill will. Call this the *Quality of Will Condition*. Ill will is, roughly, an objectionable lack of concern for morality or the morally significant interests of others. Exactly what ill will consists in is controversial. All hands agree that ill will need not involve malice, so the term is something of a misnomer: it would be better to speak of an “insufficiently good will” or perhaps an “objectionable pattern of concern.” However, nothing I say in this paper turns on any particular view of ill will, other than that ill will is a matter of what the agent wants or cares about.

On one prominent conception, an agent’s quality of will is a matter of her responsiveness to moral reasons.¹ A goodwilled person has a final desire to take courses of action that have right-making features (like kindness or fairness), and a final desire to avoid actions with wrong-making features (like cruelty). In Arpaly’s words: “To say that a person acts out of moral concern is to say that a person acts out of an intrinsic (noninstrumental) desire to follow (that which in fact is) morality.”² And a person acts from ill will if she is either insufficiently responsive to moral reasons or responsive to sinister reasons—“reasons which conflict with morality.”³ A person who tortures a puppy for fun expresses ill will because she desires the puppy’s pain, and is thus responsive to a consideration that makes the act wrong. A person who fails to aid someone who is suffering might express ill will not because she is responsive to a sinister reason but because she is insufficiently responsive to the moral consideration at hand—the person’s suffering.

Proponents of the *Quality of Will Condition* sometimes argue that it is implicit in our ordinary practice of blaming. Strawson famously claimed that the

1 Arpaly, *Unprincipled Virtue*.

2 Arpaly, *Unprincipled Virtue*, 84.

3 Arpaly, *Unprincipled Virtue*, 79.

reactive attitudes—resentment, indignation, gratitude, and the like—are always responses to the good or ill will displayed by others:

The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves; or at least on the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard.⁴

Indeed, the Quality of Will Condition appears to explain central features of our practice, most notably the range of excuses we accept as grounds for withholding blame for wrongful action. It is a central feature of our blaming practice that we withhold blame when we take the agent to have acted under duress or from certain forms of blameless ignorance. Any theory of blameworthiness must explain why these excuses work when they do, and for the Strawsonian the explanation is straightforward: these excuses negate blameworthiness (when they do) by indicating that the violation was not sourced in any improper concern for the rights or interests of others. It is important to note that the Quality of Will Condition only specifies that ill will is a *necessary* condition of blameworthiness. It is consistent with thinking there are other requirements, e.g., that the agent acted wrongly, or that she possessed the capacity to recognize and respond to moral reasons.⁵

I will first argue, against this near consensus, that the Quality of Will Condition is in fact false to ordinary moral practice.⁶ This is because ordinary practice licenses blame for agents who act wrongly from epistemically unreasonable ignorance of the wrong-making features of their act, even when the act does not express ill will. This should be especially concerning to Strawsonians, who think the norms implicit in our practice are the inevitable starting point for discerning the real norms of blameworthiness.

Second, I provide a theory of culpable ignorance (and blameworthiness more generally) on which ill will is not necessary for blameworthiness. The theory is an answer to the question: In what might the culpability of unreasonable agents be grounded, if it is not ill will? There is already an answer to this question

4 Strawson, *Freedom and Resentment and Other Essays*, 347.

5 Proponents of the Quality of Will Condition include Strawson, *Freedom and Resentment and Other Essays*; Arpaly, *Unprincipled Virtue*; Rosen, “The Alethic Conception of Moral Responsibility”; Wallace, *Responsibility and the Moral Sentiments*; McKenna, “Directed Blame and Conversation”; and Graham, “A Sketch of a Theory of Moral Blameworthiness.”

6 Baron is a fellow traveler in this position in “Culpability, Excuse, and the ‘Ill Will’ Condition.”

in the literature: capacitarianism, which says that an agent is culpable for fully unwitting wrongdoing insofar as she had, but failed to exercise, the *capacity* to be aware of the wrongness or wrong-making features of her act. But all capacitarians need to distinguish between capacities that are relevant to responsibility and capacities that are not, as will become clear. The present approach draws this line in a principled way.

My theory, which I will call *Rational Capacitarianism*, entails that when an agent acts wrongly from factual ignorance, she is blameworthy only if her ignorance is caused by a failure to exercise a *rational capacity*—a capacity to recognize reasons for believing and desiring and to be generally responsive in one's beliefs and desires to one's assessment of reasons. I argue that failures of rational capacities are relevant because only such failures can be attributed to the agent in a distinctive way. My theory shows that we do not need to choose between a quality of will view (which seems too lenient) or a standard capacitarian view (which seems too harsh).

A few preliminaries: I will consider an agent to be blameworthy for an act if she is liable to certain “hot” emotional responses for it, like resentment and indignation. This follows Strawson in reducing questions of blameworthiness to questions about the appropriateness of the negative reactive attitudes.

I will deviate from the literature in focusing on ignorance that results from epistemic unreasonableness, rather than lapses of memory or inattention. The literature on negligence tends to spotlight forgetful or inattentive agents, e.g., a parent who forgets about a child in a hot car, or a driver who does not check the rearview mirror and backs into a pedestrian. But I think it is a mistake for ill-will skeptics to emphasize these cases, because it is actually *less* apparent that forgetful or inattentive agents are culpable than it is that *unreasonable* agents are culpable. This will become clearer when I explicate the notion of a rational capacity.

1. EPISTEMICALLY UNREASONABLE IGNORANCE

There are two relevant kinds of ignorance: factual and moral. A factually ignorant agent is unaware that her act has certain empirical features that make it wrong (like being risky or harmful); a morally ignorant agent may be aware of these features, but ignorant that her act is wrong (or cruel or unfair). I will focus on factual ignorance rather than moral ignorance, because it is more difficult to see how an agent whose pertinent ignorance derives entirely from factual ignorance might be culpable.⁷ Factual ignorance is often a complete excuse. Suppose *X*

7 There is an obvious way in which morally ignorant agents can still express ill will, by being insufficiently responsive to the reasons that *make* their act wrong. A Nazi who tortures

poisons *Y* by putting a substance in her coffee. If *X* innocently believed that the substance was sugar (perhaps because it was mislabeled, and she had no reason to doubt the labeling), then *X* is excused—she is not blameworthy for the wrong she committed.⁸ Strawsonians explain this in a straightforward way: if *X* was ignorant of the harm she would cause and acted only because she was ignorant in this way, the wrong was not sourced in any improper concern.

Factual ignorance is only an excuse, of course, when it is not the foreseeable upshot of some prior blameworthy act that expressed improper concern.⁹ If to relieve the boredom of a long drive I ingest a mind-altering drug and hit a pedestrian whom I do not notice crossing the street, it is no excuse that I did not know he was there. Since I am blameworthy for recklessly taking the drug, I am blameworthy for my ignorance.

Ignorance may also not excuse if it is *motivated* by ill will, as when a racist's racial animus leads him to falsely believe a Black candidate is unqualified for a job. If an agent acts from motivated ignorance, she acts from improper concern, since her ignorance expresses improper concern. Call ignorance that satisfies neither of these conditions—that is neither motivated by ill will nor the upshot of prior blameworthy conduct—*morally blameless* ignorance (leaving open that it may be criticizable in some other way, e.g., epistemically criticizable). In what follows I set aside morally blameworthy ignorance. The plan is to focus on *blameless ignorance* and to ask when and why it amounts to an excuse.

Most theorists hold that acts performed from blameless factual ignorance are always blameless. For instance, Rosen says:

If [the agent] has been neither negligent nor reckless in the management of his opinion, then his ignorance is blameless and so is the act done from ignorance.¹⁰

innocent prisoners and is fully convinced he is doing the right thing is still insensitive to the suffering of the prisoners. This is why many quality of will theorists do not think *moral* ignorance is exculpating.

- 8 In “Does Moral Ignorance Exculpate?” Harman points out that it is not ignorance that is potentially exculpating, but false belief. If you have a 0.5 credence that a substance is poison and a 0.5 credence it is sugar, then you are ignorant that it is poison, but this does not get you off the hook for poisoning someone. You are only off the hook if you have the false belief that the substance is *not* poison. This point is well taken. “Ignorance” of *p* in this paper should be construed as shorthand for false belief that not-*p*; further qualifications may also be needed. For discussion, see Harman, “Does Moral Ignorance Exculpate?”
- 9 Rosen, “Culpability.”
- 10 Rosen, “Culpability,” 63. By “neither negligent nor reckless,” Rosen means that the person's ignorance is not the upshot of a previous negligent or reckless *act*. It does not cover pure

And Zimmerman says, “culpability for ignorant behavior must be rooted in culpability that involves no ignorance.”¹¹ Indeed, it may seem almost a truism that an agent is not blameworthy for a wrong act if she was blamelessly unaware of its wrong-making features. Consider this case of Arpaly’s: Boko Fittleworth (a character in a P. G. Wodehouse novel) captures a man he finds hiding in his future father-in-law’s garden shed, thinking he is a burglar.¹² In fact, he is not a burglar; his presence is part of a secret plot in which the future in-law is willingly participating. Boko is, obviously, completely blameless for this honest mistake.

But Rosen’s view is not as obvious as it may first appear. We must distinguish *epistemically reasonable* from *epistemically unreasonable* ignorance. The cases that support the view that blameless ignorance is always exculpating tend to be cases of epistemically justified ignorance: cases in which the agent drew sound conclusions based on the evidence available to her. I am clearly off the hook for poisoning you if my belief that the substance was sugar was not only unsourced in ill will, but completely reasonable (for instance, because the canister was labeled “sugar” and I had no reason to think this was misleading). But what if the false belief from which an agent acts is, though blameless, highly irrational? Suppose the canister was labeled “cyanide—not for consumption!” but I unreasonably, though innocently, assumed it was a joke. Am I still off the hook?

Now, irrationality is sometimes the result of underlying bad will. I wish to set these cases aside (since they are instances of *motivated* ignorance), but since it may be hard to distinguish them from cases of innocent irrationality, it is worth saying a bit more about them. In *Unprincipled Virtue*, Arpaly asks us to consider someone who claims “I hate him because he is disgusting.” We are likely to infer that the speaker “puts the cart before the horse”: it is the hate that motivates the belief that the person is disgusting, not vice versa.¹³ Consider a young academic who irrationally believes that women are awful at abstract thought, despite being surrounded by brilliant female colleagues. Again, we would reasonably surmise that the man’s irrational belief is rooted in some unsavory motive, perhaps a discomfort with women’s presence in academia. These cases of *motivated* irrationality or *motivated* ignorance are clearly possible. And when an agent acts from motivated ignorance, she acts from ill will in a straightforward sense, because her ignorance itself is the result of a bad desire.

The cases that interest me are not like this. They are ones in which the agent’s

epistemic irrationality, which will be my focus. Rosen’s view is actually more extreme, as he thinks that blameless moral ignorance is also always excusing. But we can set this aside.

11 Zimmerman, “Moral Responsibility and Ignorance,” 417.

12 Arpaly, *Unprincipled Virtue*.

13 Arpaly, *Unprincipled Virtue*, 105.

irrationality is unmotivated. It is clear that human beings are sometimes just poor thinkers. As Kahneman and Tversky have shown, we are subject to a range of cognitive biases.¹⁴ For instance, we tend to be bad at reasoning about probabilities because we neglect base rates and rely too heavily on examples that immediately come to mind (the “availability heuristic”). And sometimes, of course, people simply form beliefs that are incongruent with the overall balance of their evidence, even when they are not affected by a particular bias.

Given the prevalence of garden-variety irrationality, it is clearly *possible* that someone might act badly from ignorance that reflects unmotivated irrationality but that is not further sourced in some improper pattern of concern. Whether or not such cases are common is, of course, another matter. What I hope the reader will appreciate at this point is that such cases are *possible*, and that the right theory of blameworthiness should properly classify them. Are agents who act wrongly from *unmotivated* irrational ignorance culpable?

Of course, it can be very hard to tell in any given case whether the agent’s irrational ignorance is badly motivated or not. That is why, in constructing cases for the purposes of inspecting our sense of culpability, we should make it as clear and vivid as possible that the agent’s pattern of concern is entirely appropriate. But some help from the reader cannot hurt: I ask the reader to imagine the characters as being as goodwilled as they could possibly be—as the sort of people who really are concerned with morality and others’ interests as much as they should be, but who suffer from bouts of ordinary unmotivated irrationality.

2. SOME CASES

Sher discusses cases of “poor judgment” for which we are likely to blame the agent.¹⁵ In *Home for the Holidays*, Joilet, a homeowner afraid of burglars, hears movement in the kitchen. She grabs her gun, runs downstairs, and shoots the intruder, only to discover that it was her son who came home early for the holidays. In *Colicky Baby*, Scout gives a crying baby a mixture of vodka and juice to ease its digestive pains.

In each of these cases, it is possible—in fact, easy—to interpret the agent as acting from improper concern. We suspect that Scout knows, at least at some level, that giving a baby vodka is dangerous but just does not care enough. It is also *possible* that the agent is properly concerned but merely foolish, which is the interpretation Sher intends. Perhaps Scout would *never* give a baby something she thought was dangerous, but received some bad advice about how to

14 Kahneman and Tversky, “Judgment Under Uncertainty.”

15 Sher, *Who Knew?* 26.

calm crying babies that she credulously accepted, and now she sincerely believes that giving a baby vodka is normal and safe. Sher contends that these agents are blameworthy, despite their unawareness, and I agree.

Given that it is so tempting to read improper concern into these agents, I think it is helpful to focus on a more detailed case. Consider the following:

Bridge Collapse: Marie is in charge of conducting a stress test on a local bridge, which involves placing weights on the span to see how it responds. Marie has the option of closing the bridge during the test, which is costly and inconvenient, but sometimes required as a precaution. Whether to close the bridge is a matter of judgment reserved for the engineer; the engineer is charged with determining whether the risk of harm is “significant” and then determining whether it should be closed based on this classification. Marie is an experienced and responsible engineer, who always closes the bridge when she believes a stress test is too risky, and usually makes the right call.

But as Marie deliberates on this occasion, she commits a serious error of professional judgment. Given the magnitude of the stress test being conducted and the age of the bridge, it would in fact be *highly* imprudent not to close it. Marie ought to know this, given her training and the evidence (which is extremely clear). She should conclude that the risk of collapse is significant and bring to bear her knowledge that in such cases the bridge should be closed. She is fully capable of drawing this inference; but she does not.

Tragically, the stress test places too much weight on the bridge span and the bridge collapses, violently killing a pedestrian. Marie blames herself for the death, as does the public, once they realize it was the result of her serious error. It is wholly clear to Marie after the collapse that her decision to keep the bridge open was highly irrational, a textbook case of poor judgment on the part of an engineer.

Marie is not improperly concerned. For Marie to have acted from improper concern, her failing to close the bridge must have expressed either indifference to moral considerations or responsiveness to sinister considerations. The relevant moral consideration is the danger that the stress test poses to bridge users. However, Marie is not indifferent to the risk she imposes on others; she always closes the bridge when she believes a stress test is too risky. Her pattern of concern is impeccable in this respect. The problem is that she is unaware, on this occasion, that the stress test is too risky, so she does not know to close the bridge.

Might her ignorance express improper concern? It would if it were motivated

by, say, a desire to avoid having to close the bridge on this occasion (perhaps out of laziness). If it were, Marie's ignorance would express a pattern of improper concern consisting of her overvaluing her own convenience, even at the cost of serious risk to others.¹⁶ But Marie's ignorance was simply the result of a concern-neutral cognitive error, not improper concern of this sort.

If one is still tempted to posit improper concern to explain Marie's error, it may be helpful to think of her ignorance as produced by the operation of a cognitive bias (since this can positively replace an explanation that appeals to insufficient concern). Perhaps it is a bright sunny day and Marie has difficulty imagining the possibility of a disaster on such a lovely day. People tend to underestimate the probability of bad outcomes on sunny days because bad outcomes are less likely to come to mind on sunny days, and thanks to the *availability heuristic*, people often estimate the probability of an event based on the ease with which similar events come to mind.¹⁷ Since Marie's ignorance can be fully explained by the operation of this "cold" bias, there is certainly no need to posit improper concern to explain why Marie neither knows nor suspects the stress test is risky.

Perhaps there is another way Marie could be improperly concerned: maybe Marie just does not care enough about the rationality of her own beliefs. Rosen observes that "we are under an array of standing obligations to inform ourselves about matters relevant to the moral permissibility of our conduct: to look around, to reflect, to seek advice, and so on."¹⁸ We are not *always* morally obligated to check ourselves for bias; I do nothing morally wrong when I complacently believe that Linda is more likely to be a feminist and a bank teller than just a bank teller, or idly question whether the moon landing really took place. But there are cases in which one plausibly should check oneself for irrationality, because the stakes are high and the cost of checking oneself is low. Maybe Marie displays improper concern by not taking more measures to ensure that her judgment is sound, for instance, by seeking advice.

But it is not plausible that Marie displayed improper concern by not seeking advice or rechecking her belief, because, as stipulated, whether to close the bridge is a matter of judgment reserved to the engineer according to the opera-

16 Nor is Marie's ignorance the result of an epistemic vice. FitzPatrick argues that we can be responsible for ignorance that was the result of "voluntary exercise of vices," such as "overconfidence, arrogance, dismissiveness, laziness, dogmatism, incuriosity, self-indulgence, contempt, and so on" ("Moral Responsibility and Normative Ignorance," 605); this is not one of these cases.

17 It is not that the bias renders Marie incapable of getting the right answer; she is a competent engineer and so fully capable of making the right classification. The sunny day makes the problem slightly harder, but not beyond her power by any means.

18 Rosen, "Culpability," 63.

tive procedural norms. The engineer is given latitude to determine whether the risk of harm is “significant,” so Marie is entirely within protocol when she makes the call based on her own deliberation. It is not plausible that Marie displays insufficient concern for others’ safety by failing to go *above and beyond* the protocol, which she has no reason to suspect is flawed. We can even imagine that Marie rechecks her informal calculation, but is subject to the same bias each time, so continues to underestimate the danger posed by the stress test.

There is thus no reason to think that Marie’s pattern of concern is defective (and in any case, we can just stipulate that this is the case). Now the critical question: Are Marie and other comparably irrational agents blameworthy for their foolish decisions? I believe they are. Consider the sort of excuses they might make. Scout in Colicky Baby might sincerely plead that she did not know that vodka was dangerous, that she meant well. Marie might maintain that she followed proper protocol, that she just made an honest mistake. These assertions could all be true, and they justify a certain amount of sympathy for Scout and Marie. But I am inclined to think that they are only partial excuses, even though they establish the goodwill of the agents. Marie’s foolishness *killed* a person. Scout gave a baby alcohol poisoning. These agents were not *incapable* of exercising better judgment. Marie had never made an incorrect call before, and while a sunny day might affect her judgment, nobody can claim she rationally *debilitated* it and that that made her incapable of reasoning correctly. Scout is a normal adult perfectly capable of recognizing that “just give the baby vodka” is poor advice. Our complaint against these agents is simply that they could have and should have known better.

To be clear, my claim is not just that these agents are liable to criticism or that they should compensate the victims if they can. I think they are fully blameworthy, in the sense of meriting “angry blame”: resentment and indignation. To be sure, they merit much less blame than agents who perform comparable acts intentionally. But they are culpable nonetheless.¹⁹

19 It is also worth noting that Anglo American criminal law allows liability in some cases of unreasonable ignorance without requiring any further finding of ill will. A successful mistake-of-fact defense requires that the defendant demonstrate his mistake was reasonable: one that an ordinary person would (or might) have made under the circumstances. In *Commonwealth v. Pierce* (138 Mass. 165 [1884]), a leading case on ignorance of fact, the defendant was a doctor who attempted to cure a patient by wrapping her in kerosene-soaked clothes. Though the doctor was determined to be acting without knowledge of the risk, he was convicted of involuntary manslaughter on the ground that his ignorance was unreasonable. Such agents may be less culpable than people who cause harm knowingly or recklessly. But it is simply not part of law to excuse them altogether on the ground that they were irrational or stupid or otherwise unreasonable rather than ill willed.

One might worry that, despite every stipulation to the contrary, our sense of culpability is influenced by an inclination to attribute ill will or “mismanagement of opinion” to these agents. Rosen argues that though we are sometimes inclined to blame agents who act wrongly from blameless ignorance, this is due to a failure to hold clearly in mind all the relevant facts. When we *do* manage to bear in mind that the agent’s ignorance was blameless, our resentment evaporates, as Rosen claims in this discussion of our reaction to an ignorant ancient slaveholder (who is unaware that slavery is morally wrong):

We are no doubt powerfully inclined to blame [the ancient slaveholder]—so long as we ignore the stipulated fact that he is blameless for not knowing that his slave deserves much more. When we bear this in mind—when we “zoom out,” as it were—then (I claim) our sense of his culpability evaporates. It is as if in blaming him we are thinking that he should have known better; he should have known that his action expresses an attitude that would merit intense resentment; he should have known that his act is the sort of act for which one should feel shame in retrospect. When we remind ourselves what it would have taken for him to know these things, and in particular that it’s not his fault that he doesn’t know them, then our resentment is properly blocked.²⁰

So, for instance, if we manage to hold in mind that Marie followed proper protocol—and hence, her ignorance was blameless—our sense of culpability should evaporate according to Rosen. But I do not find Rosen’s claim persuasive here. Rosen claims that the “should” thoughts that sustain resentment are falsified when an agent’s ignorance is blameless: “he should have known better,” “he should have known that his action expresses an attitude that would merit intense resentment,” and so on. These thoughts are indeed false if the “should” is interpreted (as it is by Rosen) as indicating the agent should have *done* something differently in the management of her opinion.²¹ But this is not the only way to interpret the relevant “should” thoughts that trigger resentment. If the agent’s ignorance is unreasonable, then a different, epistemic version of the “should” thought is available: “*X* *rationally* should have known better.” Her epistemic performance fell below some operative epistemic standard. I maintain that

20 Rosen, “Culpability,” 73.

21 According to Rosen, someone is blameworthy in her ignorance only if it can be traced to some prior blameworthy act or omission, like a failure to satisfy one’s procedural-epistemic obligations. Rosen stipulates that the slaveholder complied with his procedural-epistemic obligations and hence his ignorance is blameless; “we are under no obligation to rethink the uncontroversial normative principles that form the framework for social life” (“Culpability,” 65).

thoughts about the agent's epistemic failings suffice to sustain resentment and indignation in these cases. These emotions are stoked simply by the thought that the agent could have and should have (according to some operative epistemic standard) known better, and this is so even if we expressly cancel the thought that the agent's failure of rationality is fueled by ill will on her part. We imagine the agent having all the evidence in front of her, in full possession of her cognitive faculties, yet drawing an utterly unreasonable conclusion, to dire effect. This is enough to fuel resentment's fire.

Of course, just because we *tend* to resent agents who exercise poor judgment does not mean we are *correct* to resent them. One might dismiss this inclination as seriously misguided, and call for major revision to our ordinary practice of blame and punishment. But I do want to point out that this is in some tension with the Strawsonian approach to moral responsibility that originally motivated the ill-will condition. Strawson believed that any credible account of the conditions of responsibility must vindicate ordinary practice at least to a significant degree: our practice of holding one another responsible is an "essential feature of our way of life" and therefore "neither calls for nor permits, an external 'rational' justification."²² Strawson thought ill will was necessary for blameworthiness because this is what ordinary practice reflects: "The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves."²³ But I have claimed that ordinary practice with the reactive emotions does not always reflect a demand for goodwill: it sometimes reflects a demand for *reasonableness*, as when we resent agents for their poor judgment. If that is right, then Strawsonians who consider the ill-will condition to be vindicated by the standards of ordinary practice have a predicament.

To see why Strawsonians of this sort are in trouble, consider Rosen's Strawsonian view. Rosen is a Strawsonian in the sense that he both (1) endorses the ill-will condition and (2) believes this is vindicated by an aspect of ordinary practice, "what the personal reactive attitudes rest on."²⁴ According to Rosen, resentment does not only *depend* on the belief that the agent expressed ill will, but *constitutively involves* this thought. That is, just like fear of *X* contains the thought "*X* is dangerous," resenting *X* for *A* contains the thought that "*X*'s doing *A* expressed an objectionable pattern of concern."²⁵ And this is why, for Rosen, agents are blameworthy when they express ill will: expressing ill will makes or-

22 Strawson, *Freedom and Resentment and Other Essays*, 23.

23 Strawson, *Freedom and Resentment and Other Essays*, 17.

24 Strawson, *Freedom and Resentment and Other Essays*, 347.

25 Rosen, "The Alethic Conception of Moral Responsibility," 14.

dinary resentment appropriate, given its constituent thought. (Just like fear of X is appropriate just in case X is *in fact* dangerous.)

But is this right? On a view like Rosen's, we can get a handle on the thoughts implicit in an emotion by asking which sort of discoveries cause it to evaporate. For instance, if fear usually evaporates upon learning or perceiving that X is not dangerous, this is evidence that "X is dangerous" is a constituent thought of fear. Rosen contends that resentment always evaporates when we "zoom out" and bear in mind the agent's ignorance was blameless. But I have argued that the thought "X could have and should have known better" is enough to sustain resentment, even in the absence of ill will. If that is right, this suggests the ill-will thought is not constitutive of resentment.²⁶ And this is trouble for the foundation of Rosen's account, since it removes the very justification for the ill-will condition.

I concede that these arguments are not decisive. Whether we regard unreasonable ignorance as sufficient for blameworthiness should ultimately depend on the overall plausibility of a theory that rejects the quality of will constraint and provides alternative conditions of blameworthiness. My aim in what follows is to sketch the beginnings of such a theory.

3. RECOVERING THE VIEW OF BLAMEWORTHINESS

Ill-will theorists have a pleasingly simple theory of blameworthiness: X is blameworthy for a wrong act A iff in doing A , X expressed ill will (and, perhaps, possessed a certain requisite capacity, like the capacity to assess and respond to moral reasons—what Strawson called a "moral sense").²⁷ This theory must be revised if we are to account for the culpability of unreasonable but non-ill-willed

²⁶ One might object to this argument on grounds that it is clearly possible to experience a reactive emotion inappropriately. Fear contains an "X is dangerous" thought, but someone can recoil in fear at the sight of what she knows to be a harmless snake. Might resentment for unreasonableness be like this—possible and understandable but not appropriate? It is worth noting that even when one knows that a snake is harmless, it can still give the *appearance* of being dangerous—we are hardwired to see the serpentine silhouette as indicating "danger." Hence one might have a quasi-perceptual "X is dangerous" thought, in spite of one's explicit knowledge. But we resent for unreasonableness even when it is *manifestly clear* that the agent was not ill willed (as in Bridge Collapse). Though it is always *possible* that our resentment is based on some illusion of ill will, the best explanation for this, I believe, is simply that we can resent without thinking anything like the ill-will thought. As Rosen acknowledges: "An account of the thoughts implicit in resentment is unacceptable if it is clear on reflection that we can resent X for A without thinking the thoughts posited by the account" ("The Alethic Conception of Moral Responsibility," 75).

²⁷ Strawson, *Freedom and Resentment and Other Essays*.

agents. How might we do so? One possibility is to resign ourselves to an unappealing disjunctive account: X is blameworthy for a wrong act A iff in doing A , X expressed objectionable will, *or* unreasonable ignorance (perhaps, again, against the backdrop of possessing the capacity to assess and respond to reasons). This would be to concede that sometimes we resent agents in virtue of ill will, and sometimes unreasonable ignorance, and that no more unified account of our practice can be given.

But such resignation is unnecessary. There is, I believe, a unified view of moral blameworthiness consistent with our practice that accounts for the blameworthiness of agents who act from unreasonable ignorance, and ill will (though the blameworthiness of ill-willed agents will not lie in their ill will, *per se*, but in their unreasonableness). On this account, agents are blameworthy whenever they act wrongly from an unreasonable attitude—a belief, intention, or desire—that is attributable to them, in a sense I will explain. Acting from unreasonable ignorance and acting from ill will are both ways of acting from unreasonable attitudes attributable to the agent.

To see how cases of blameworthy ill will and blameworthy ignorance might be unified, let us observe that, in both sorts of cases, the agent's wrong act reveals an unreasonable attitude (belief, intention, or desire): one that is held on the basis of bad reasons, or is insufficiently sensitive to the reasons the agent has. This assumes that there are rational norms on final desires as well as beliefs, something that Humeans have long denied. But let us suppose that final desires are rationally criticizable, so that, for instance, a person who finally desires to inflict suffering on others wants something she has no good reason to want. Then wrong action from bad desire and wrong action from irrational ignorance both express unreasonableness, as can be seen in the following pair of cases:

Ill Will: Suppose Smith kills Jones out of malice. There are conclusive reasons for Smith not to kill Jones (just as there are conclusive reasons for anyone not to kill gratuitously), but Smith is irrationally insensitive to these reasons. Hence the *intention* or *desire* that moves Smith to act is unreasonable.

Unreasonable Ignorance: Suppose Smith shoots and kills Jones while horsing around due to a false belief that playing around with a gun is not risky. Smith has conclusive reason to believe that horsing around with a gun is risky, but is irrationally insensitive to these reasons. Hence the *belief* from which Smith acts is unreasonable.

Of course, the sort of reasons the agent improperly assesses in each case are dif-

ferent. In Ill Will, the relevant reasons are reasons for intending or desiring. In Unreasonable Ignorance, the relevant reasons are reasons for belief. But insofar as each is subject to rational norms, both agents act from attitudes that are rationally criticizable.

To be sure, in Ill Will Smith also expresses ill will in the ordinary sense; he has a final desire to do what is in fact wrong. A traditional Strawsonian will say that this bad desire suffices to make him blameworthy, regardless of whether it is rationally faulty. What I am considering is that it is not the badness of the desire that renders Smith culpable, but its unreasonableness.

Michael Smith argues for something like this view: it is irrationality in either an agent's beliefs or desires that renders her "at fault" for wrong conduct.²⁸ First, consider Gary Watson's contrast between self-indulgence, weakness, and compulsion, to which Smith appeals:

Suppose that a particular woman intentionally takes a drink. To provide an evaluative context, suppose she ought not to have another because she will then be unfit to fulfill some of her obligations. Preanalytically, most of us would insist on the possibility and significance of the following three descriptions of the case: (1) the reckless or self-indulgent case; (2) the weak case; and (3) the compulsive case. In (1), the woman knows what she is doing but accepts the consequences. Her choice is to get drunk or risk getting drunk. She acts in accordance with her judgment. In (2) the woman knowingly takes the drink contrary to her (conscious) better judgment; the explanation for this lack of self-control is that she is weak-willed. In (3), she knowingly takes the drink contrary to her better judgment, but she is the victim of a compulsive (irresistible) desire to drink.²⁹

Smith contends that we blame both the self-indulgent woman and the reckless woman. We blame the self-indulgent woman for "having the wrong belief about what she should do in the circumstance" when she could and should have known better. We blame the weak woman, "[not] for her belief—she has the belief she should have, after all—but rather for her failure to act on that belief" when she could have.³⁰ By contrast, we do not blame the woman who acted compulsively at all, because she believed correctly and could not have done otherwise than she did.

Having the wrong belief (like the self-indulgent woman) and having the wrong intention or desire (like the weak woman) are both forms of irrational-

28 Smith "Rational Capacities."

29 Watson, "Skepticism about Weakness of Will," 324.

30 Smith, "Rational Capacities," 18.

ity. The self-indulgent woman fails to recognize the reasons against intending/desiring to take the drink; the weak woman recognizes these reasons, but fails to desire in accord with them. Says Smith:

If people possess, but fail to exercise the relevant [rational] capacities, then there is something that they could have done in a perfectly mundane sense: they could have exercised the capacity to access the available evidence, or to believe in accordance with their evidence, or to . . . desire, in the way that's rationally required. Their failure to access or believe or . . . desire correctly is therefore their fault . . . and so too is their failure to perform the act that they would otherwise have performed.³¹

As Smith is careful to emphasize, agents must have the *capacity* to assess reasons, and to modulate their beliefs and desires accordingly, in order to be held responsible for their unreasonable conduct. Nearly everyone agrees that very young children, animals, and the psychotic are not liable to blame even when their acts are sourced in unreasonable intentions, desires, or beliefs. This is presumably because they substantially lack the capacity to assess reasons and to believe and desire in line with assessments of reasons, so their failures to do so are not their fault.

In saying their unreasonableness is not their “fault,” we are not yet making a point about moral blameworthiness or the lack thereof in the sense at issue in this paper. As I read him, Smith’s talk of “fault” in this context concerns a different notion: whether the failure is properly attributable to the agent, i.e., whether she is properly criticized *qua* agent on the basis of it. If an agent possesses the relevant capacity, then—assuming the capacity is not masked (because she has been drugged or is excessively tired or whatnot)—failures to properly exercise the capacity are “down to her.” But if she lacks the capacity, they are not “down to her” (they are, rather, attributable to the absence of the capacity), and she is not criticizable for these rational failings.

So, having the relevant capacity—that is, the capacity to assess reasons and modulate one’s attitudes in light of these assessments—renders one responsible, *in the attributability sense*, for one’s attitudes. That is Smith’s point to which I add: agents are blameworthy *in the resentment-involving sense* when they act from unreasonable attitudes that are attributable to them.

Taking account of all this, the resultant theory of blameworthiness, which I will call *Rational Capacitarianism*, is as follows:

31 Smith, “A Constitutivist Theory of Reasons,” 11.

Conditions of Blameworthiness: X is blameworthy for a wrong act A iff X 's doing A expressed an unreasonable attitude that is attributable to her.³²

Theory of Attributability: An attitude is attributable to X iff X has the capacity to assess the reasons for and against the attitude and modulate her attitude in light of this assessment during the period in which the attitude was formed and sustained.³³

Take again a case of blameworthy ill will: Smith kills Jones out of malice. On Rational Capacitarianism, Smith is blameworthy because his wrongful act is rooted in some unreasonableness that is down to him:

1. What is the unreasonable attitude expressed? The intention or desire to kill Jones.
 2. Why is the intention attributable to Smith? Smith is a competent adult and had the ability to assess reasons for and against intending/desiring, and to modulate his intentions/desires accordingly when the attitude was formed and/or sustained.
 3. Why is the intention unreasonable? There are two possibilities. If Smith fails to recognize that there is conclusive reason not to kill Jones despite awareness of the pertinent facts (analogous to the self-indulgent woman), his unreasonableness lies in his flawed belief. If he recog-
- 32 The "expression" of an irrational attitude, we might say, involves its appearance in a certain standard story of action. On the standard story, an agent's ϕ -ing is an action when it results non-deviantly from her beliefs and desires combining in a certain way, i.e., when she (1) has some final desire to ψ and (2) a belief that ϕ -ing is likely to bring about her ψ -ing, which results in an instrumental desire to ϕ . An action "expresses" an irrational belief or desire when one of the beliefs or desires that produces the action in this way is irrational.
- 33 My account bears similarity to Fischer and Ravizza's reasons-responsiveness account of moral responsibility, but should be distinguished. Fischer and Ravizza provide a necessary condition for responsibility: a condition that an agent must satisfy if she is to be morally responsible for her actions. By "responsible," they simply mean an apt target of the reactive attitudes, both negative and positive: "Our Strawsonian view of moral responsibility allows for responsibility for 'morally neutral' behavior. For instance, one can be morally responsible for simply raising one's hand" (*Responsibility and Control*, 8). In contrast, I am arguing that failures of the capacity for reasons-responsiveness (in the sense I have described) are sufficient for blameworthiness when expressed in wrongful action, and moreover, failures of the capacity ground an agent's blameworthiness in cases of wrongful action from ignorance. (Fischer and Ravizza do not tailor their account to ignorance; their focus is "the freedom-relevant condition on moral responsibility, rather than the epistemic condition"; *Responsibility and Control*, 13.) Someone could accept Fischer and Ravizza's view that reasons-responsiveness is a necessary condition for responsibility without endorsing my view that failures of this capacity are sufficient for, and moreover are the ground of, blameworthiness when expressed in wrongful action.

nizes there is conclusive reason not to kill Jones but kills him anyway (analogous to the weak woman), his unreasonableness lies in his flawed intention or desire.

Now return to Marie in Bridge Collapse. Marie also acts from an unreasonable attitude that is attributable to her: the irrational belief that the stress test is not risky. Why is the attitude unreasonable? It reflects an improper assessment of the considerations that favor closing the bridge. Why is it attributable to Marie? Marie could have evaluated the evidence properly and believed in accord with it; she was not rationally debilitated. Hence the collapse is Marie's fault, and she is blameworthy for it.

It is not obvious that what makes an attitude attributable to an agent, of course, is that it expresses the failure to exercise a rational capacity of this sort, though this is what bad desires and unreasonable ignorance appear to have in common. One might alternatively maintain that an attitude is attributable to an agent only if it actually reflects her evaluative judgments. This is the view of Angela Smith:

To say that an agent is morally responsible for something ... is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it.... Most of our desires, beliefs, and other attitudes seem to meet this condition of judgment-dependence, even though they do not commonly reflect a choice or decision, and are not normally under our voluntary control.³⁴

The problem with this view is that it cannot explain responsibility in cases of weakness of will. Both the self-indulgent and the weak woman are responsible in Watson's example, but only the self-indulgent woman's act reflects her evaluative judgment. The weak woman acts *against* her better judgment, taking the drink when she knows she should not.

What unifies the class of attitudes that are attributable to us, I believe, is not that they actually reflect our evaluative judgments, but merely that they are *apt* to reflect them—they are the sorts of attitudes that *could* be brought into alignment with our evaluative judgments. These attitudes, which include beliefs, desires, and intentions, are under a kind of "rational control," in the sense that they are capable of being regulated by, hence controlled by, an agent's recognition of norms, however imperfectly.³⁵ Some writers see this sort of "control by the

34 Smith, "Control, Responsibility, and Moral Assessment," 369–70.

35 The presence of "rational control" is what leads Pettit and Smith to remark that "[An agent's] beliefs do not just come and go in a natural procession of events.... The subject is

agent's view of the reasons" as strongly analogous to voluntary control, and so as amounting to a kind of freedom akin to free will.³⁶ But that is not necessary for present purposes. All that matters is that when an attitude is formed in an agent who possesses the capacity to respond appropriately to reasons and to modulate his attitudes in light of his assessment of the reasons, that attitude is, in every morally relevant sense, his own.

4. CAPACITARIANISM

There is another account of responsibility that widens the scope of attributable attitudes. On this view, not only failures of rational capacities are attributable to the agent, but failures of any capacity, or perhaps of any "psychological capacity." Capacitarians agree with me that quality of will is not necessary for blameworthiness.³⁷ Capacitarians contend that agents with a decent quality of will are blameworthy for wrong acts performed from ignorance if, roughly, they could have and should have known that the act had the relevant features—that is, if their ignorance is the result of the improper exercise of some capacity.³⁸

Capacitarians tend to take cases of forgetful negligence as paradigm cases of culpable ignorance, which I have deliberately avoided doing. For instance, Sher focuses on a case that he calls *Hot Dog*, in which Alessandra forgets her dog Sheba in a hot car. Alessandra forgets about Sheba not because she cares too little about Sheba, but because she gets distracted by a situation occurring at her child's school. Sher contends that Alessandra is blameworthy because she

certainly not a mere passive or mechanical system. She does not just revise her beliefs and desires autonomically, or at any rate, not when they operate beyond the reach of the occasional disabling obstacles that get in her way. She revises them under the spur of recognizing what the relevant norms require of her" (*Freedom in Belief and Desire*, 442). Some have argued that a kind of capacity-relative control suffices for an agent's ignorance to itself be blameworthy (Rudy-Hiller, "A Capacitarian Account of Culpable Ignorance"). For instance, on Rudy-Hiller's capacitarian account, "an agent can be directly in control (in the responsibility-relevant sense) of his ignorance, since he can have direct capacitarian control over whether he notices, remembers, or otherwise is aware of a relevant consideration" (415). And this means an agent can be "directly responsible" for his ignorance (in the sense of being culpable for it) (407). *Contra* Rudy-Hiller, I think agents can be blameworthy for these attitudes themselves; an agent's defective attitude must be expressed in volitional action to be blameworthy.

36 See, e.g., Smith and Pettit, "Freedom in Belief and Desire."

37 Sher, *Who Knew?*; Rudy-Hiller, "A Capacitarian Account of Culpable Ignorance"; Clarke, "Negligent Action and Unwitting Omission"; Murray, "Responsibility and Vigilance"; and Amaya and Doris, "No Excuses."

38 Rudy-Hiller, "A Capacitarian Account of Culpable Ignorance," 405.

could have and should have remembered that Sheba was in the car; her failure of memory fell below “some applicable” standard that she could have met.³⁹

But we must be careful, as we are not always responsible for failing to exercise capacities we possess. Consider the following case.

Cheerleading Disaster: Sarah is a base on the cheerleading squad tasked with catching Jane, a flyer, after an air flip. Sarah is strong enough to catch Jane and has done so many times before, so she has the capacity to catch her. But on this occasion, her muscles fail her, and Jane falls through her grip, resulting in serious injury to Sarah.

The problem is not a failure of intention or effort on Sarah’s part; Sarah exerted more than enough effort as would normally be needed. Her muscles simply did not cooperate with her intention. It is true that Sarah “could have and should have” caught Jane, but this is intuitively insufficient for her moral blameworthiness when we are clear about why she did not do what she could have and should have done. The challenge for capacitarrians is to distinguish morally relevant capacities from irrelevant ones.⁴⁰

Rational capacitarrians draw the line at rational capacities; all and only failures of rational capacities (including the capacity to form evaluative judgments and modulate one’s attitudes accordingly) are attributable to the agent. The underlying idea is that an agent is responsible only for her attitudes that are apt to reflect her evaluative judgments, that *could* be brought into alignment with them, and would if we were perfectly rational—the attitudes that are under “rational control.”

This raises the question: Is what we remember under rational control? Are failures of memory attributable to us; are we answerable for them? This is a difficult question. To be sure, what we remember is sensitive to some degree to our judgments about reasons. Suppose at 8 AM I think to myself that I really ought to pick up some groceries after work. This makes me more likely to remember to pick up groceries after work. If one judges something to be very important, this increases the chance that thoughts about it will come to mind at the right time, that one’s memory will be jogged by things pertaining to it, and so on. If a failure of memory is traceable to a failure to recognize the strength of the reasons

39 Sher, *Who Knew?* 88.

40 Sher contends that for an agent to be responsible for a failure of a capacity the failure must be “caused by the interaction of some combination of his constitutive attitudes, dispositions, and traits” (*Who Knew?* 88). I am skeptical that Sher’s proposal can explain all the cases. Marie’s failure to draw the correct conclusion in Bridge Collapse is the result of her being subject to certain biases; and being subject to these does not seem constitutive of who she is.

for carrying out the task, it is traceable to a failure of a rational capacity and the agent is responsible for it.

The harder case is when an agent recognizes the importance of performing the task, resolves to do it, and then simply forgets to bring about the planned action. This sort of mnemonic failure seems relevantly different from a failure to intend or believe in accordance with one's reasons. With belief and intention, rationality concerns roughly the synchronic harmony between one's assessments of reasons and one's first-order attitudes: ideally rational agents bring them into synchronic consistency. When a rational agent is aware of a reason, forms a judgment about what reason requires, and adjusts her first-order attitudes so as to bring them into conformity with this judgment, the episode, though not strictly voluntary, does seem to involve a kind of activity and is properly read as the agent's making the resulting attitudes her own.

Remembering, on the other hand, does not seem to involve a kind of rational activity. Resolving to remember seems more like setting an alarm clock and hoping it goes off at the right time; the episode of recollection itself either happens or not, though not in a way traceable to the self. Failures of memory seem more like "glitches" analogous to the failure of strength in *Cheerleading Disaster* than defects in the agent. Consider that, in forgetting to perform some task, one's beliefs about the reasons to perform the task are not occurrent, so remembering is not merely a matter of bringing one's assessments of reasons and one's first-order attitudes into synchronic consistency. It seems to me, therefore, that a failure of memory of this sort is not a failure of a rational capacity in the relevant sense, so is not attributable to the agent.

This may be controversial; to the extent that memory is under rational control in the sense in which beliefs, desires, and intentions are, it may make sense to blame agents for failures of memory. But precisely because this is unclear, I think it is a mistake for ill-will skeptics to emphasize forgetting cases. When we focus on cases of glitchy memory in goodwilled agents and stipulate beyond all doubt that at every stage along the way they responded appropriately to the reasons of which they were aware, exercising impeccable rational control, it is hard to see the agent as anything but the passive victim of a glitchy sub-personal module on which he was fully entitled to rely. Cases of this sort are at any rate certainly not clear cases of responsibility. Cases of rational failure on the part of competent agents, I believe, are much clearer. This theory is, of course, only a sketch; but it shows how blame for "ill will" may be more continuous with blame for unreasonable ignorance than one might have initially supposed. I hope to have shown that we do not need to choose between a quality of will view (which seems too lenient) or a standard capacitarian view (which seems too harsh). Ra-

tional capacitarianism—on which only rational capacities are relevant—offers a principled alternative.

5. CONCLUSION

I have argued that agents are blameworthy when they act wrongly from unreasonable ignorance even in the absence of ill will, and provided a theory that ratifies this verdict. The theory entails that when an agent acts wrongly from factual ignorance, she is blameworthy only if her ignorance derives from a failure to exercise a rational capacity—a capacity to recognize reasons for believing and desiring and to be responsive in one's beliefs, desires, or intentions to one's assessment of reasons.

The plausibility of a theory that licenses blame for non-ill-willed unreasonableness in part depends on what blame and blameworthiness consists in. In closing I will say a few words about this issue. Like Rosen, Gibbard, and Wallace, I have fastened on a conception of blame according to which the primary forms of blame are the negative reactive emotions like resentment and indignation, and on which an agent is blameworthy for an act iff she is liable to resentment and indignation for it.⁴¹ But this is not a complete analysis, because it does not specify what being “liable” to the negative reactive attitudes amounts to. How we fill this gap is important.

Rosen's view, for instance, is that since resentment and indignation are emotions their appropriateness conditions are given by their constitutive thoughts. Emotions involve belief-like mental states; fear involves the thought that *X* is dangerous, misery the thought that things are going badly, and so on, and an emotion is appropriate just in case its constituent thoughts are true.⁴² Accordingly, for it to be appropriate to resent *X* for *A* is for the thoughts implicit in resentment to be true of *X* and *A*.⁴³

Everything I have said in this paper is consistent with this general view. If we adopt this framework, it should be read as contesting Rosen's view of the constitutive thoughts of resentment. Rosen thinks there is a “quality of will thought” contained in resentment (“in doing *A*, *X* expressed an objectionable pattern of concern”), and I disagree. My argument for this is simply that it is possible to resent *X* for an act that expresses unreasonable ignorance, but not ill will, clear-

41 Rosen, “The Alethic Conception of Moral Responsibility”; Gibbard, *Wise Choices, Apt Feelings*; and Wallace, *Responsibility and the Moral Sentiments*.

42 Rosen, “The Alethic Conception of Moral Responsibility,” 71–72.

43 Rosen, “The Alethic Conception of Moral Responsibility,” 72.

headedly, in full awareness of this fact. Resentment is stoked in these cases by the thought that *X* could have and (rationally) should have known better.⁴⁴

If we reject Rosen's account of what it is to be "liable" to the reactive emotions, it may be harder to see how resentment directed at unreasonableness could be appropriate in the absence of ill will. The "fairness view" specifies that *X* is liable to resentment for *A* just in case it would be morally fair to resent her for *A*, and the "fittingness view" appeals to a primitive, *sui generis* relation of fittingness.⁴⁵ But how could it be fair, or fitting, to resent—which involves the withdrawal of some amount of goodwill—if there was no comparable ill will on the part of the target of the resentment?

I do not have a complete answer to this question, but I will outline the beginnings of a response. The fairness of resentment plausibly depends on whether the target of the resentment had a fair opportunity to avoid being a target of resentment. Rudy-Hiller contends, for instance, that blaming for ignorance is appropriate because an agent "can have direct capacitarian control over whether he notices, remembers, or otherwise is aware of a relevant consideration," which is simply a matter of having the "requisite abilities to do these things and a fair opportunity to exercise them."⁴⁶ I differ from Rudy-Hiller in thinking that resentment is only appropriately directed at attitudes that are under the agent's *rational* control, but the spirit of my view is the same. Agents can avoid being the target of resentment by exercising their capacity to recognize reasons for believing and desiring and to be responsive in their beliefs and desires to their assessment of reasons. This is enough, it seems, for a fair opportunity to avoid.

Resentment plausibly is a sort of sanction directed at agents who we think could have done better, aimed at ensuring the agent will do better in the future. The attitudes with respect to which this corrective activity makes sense are precisely the ones that are under the agent's rational control—the attitudes that she can modulate according to her evaluative judgments. What resentment "wants," so to speak, is for the agent to recognize and revise her defective attitude (for instance, by adopting the appropriate belief or desire); this is why resentment is often not satisfied unless the agent recognizes *why* what she did was wrong. If resentment is a correctional activity of this sort, it is easier to see how it could be appropriately directed at all forms of unreasonableness, not just ill will.

44 As Rosen acknowledges, "An account of the thoughts implicit in resentment is unacceptable if it is clear on reflection that we can resent *X* for *A* without thinking the thoughts posited by the account" ("The Alethic Conception of Moral Responsibility," 75).

45 Wallace, *Responsibility and the Moral Sentiments*; see Rosen, "The Alethic Conception of Moral Responsibility," 70.

46 Rudy-Hiller, "A Capacitarian Account of Culpable Ignorance," 415, 407.

I will conclude by stressing that the justification of serious punishment is another matter. Resentment and indignation are moderate, temporary responses that tend to lessen when the agent corrects the attitudes toward which resentment is directed. Serious punishment (e.g., a restriction of an agent's liberty) of the sort that conveys serious moral condemnation is plausibly inappropriate unless the agent expressed improper concern; as the costliness to the agent of a blaming response increases, the more it seems reasonable to require some morally objectionable feature of the agent's will to be involved.⁴⁷

Princeton University
 aayars@princeton.edu

REFERENCES

- Amaya, Santiago, and John M. Doris. "No Excuses: Performance Mistakes in Morality." In *Handbook of Neuroethics*, edited by Jens Clausen and Neil Levy, 253–72. New York: Springer, 2015.
- Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press, 2002.
- Baron, Marcia. "Culpability, Excuse, and the 'Ill Will' Condition." *Aristotelian Society Supplementary Volume* 88, no. 1 (June 2014): 91–109.
- Clarke, Randolph. "Negligent Action and Unwitting Omission." *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, edited by Alfred R. Mele, 298–317. Oxford: Oxford University Press, 2014.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 2000.
- FitzPatrick, William J. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge." *Ethics* 118, no. 4 (July 2008): 589–613.
- Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press, 1992.
- Graham, Peter A. "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research* 88, no. 2 (March 2014): 388–409.
- Harman, Elizabeth. "Does Moral Ignorance Exculpate?" *Ratio* 24, no. 4 (December 2011): 443–68.

⁴⁷ Thanks to an anonymous reviewer for this point. For invaluable comments on and discussion about this article I would like to thank Gideon Rosen, Randolph Clarke, Michael Smith, Elizabeth Harman, and three anonymous reviewers for the *Journal of Ethics and Social Philosophy*.

- McKenna, Michael. "Directed Blame and Conversation." *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 119–40. Oxford: Oxford University Press, 2012.
- Murray, Samuel. "Responsibility and Vigilance." *Philosophical Studies* 174, no. 2 (2017): 507–27.
- Pettit, Philip, and Michael Smith. "Freedom in Belief and Desire." *Journal of Philosophy* 93, no. 9 (September 1996): 429–49.
- Rosen, Gideon. "The Alethic Conception of Moral Responsibility." *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–88. Oxford: Oxford University Press, 2015.
- . "Culpability and Ignorance." *Proceedings of the Aristotelian Society Volume* 103, no. 1 (June 2003): 61–84.
- Rudy-Hiller, Fernando. "A Capacitarian Account of Culpable Ignorance." *Pacific Philosophical Quarterly* 98, no. s1 (December 2017): 398–426.
- Sher, George. *Who Knew? Responsibility without Awareness*. Oxford: Oxford University Press, 2009.
- Smith, Angela M. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138, no. 3 (April 2008): 367–92.
- Smith, Michael. "A Constitutivist Theory of Reasons: Its Promise and Parts." *Law, Ethics and Philosophy* 1 (2013): 9–30.
- . "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion." In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, 17–38. Oxford: Oxford University Press, 2003.
- Strawson, P. F. *Freedom and Resentment and Other Essays*. Abingdon, UK: Routledge, 2008.
- Tversky, Amos, and Daniel Kahneman. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (September 27, 1974): 1124–31.
- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.
- Watson, Gary. "Skepticism about Weakness of Will." *Philosophical Review* 86, no. 3 (July 1977): 316–39.
- Zimmerman, Michael J. "Moral Responsibility and Ignorance." *Ethics* 107, no. 3 (April 1997): 410–26.

THE MORAL CLOSURE ARGUMENT

Matt Lutz

A SKEPTICAL HYPOTHESIS ARGUMENT introduces a scenario—a skeptical hypothesis—where our beliefs about some subject matter are systematically false, but our experiences do not discriminate between the case where our beliefs are true and the skeptical scenario where they are not. Because we are unable to rule out this scenario, we do not know that any of our beliefs about the subject matter are true. As one famous skeptical hypothesis argument goes: I cannot rule out the hypothesis that I am being deceived by a demon. Therefore, I cannot know anything about the external world. By similar token, a moral skeptical hypothesis argument is an argument that moral knowledge is impossible for agents like us in situations like ours, because we are unable to rule out some skeptical hypothesis.

In this paper, I will defend a moral skeptical hypothesis argument—the Moral Closure Argument—against a number of objections. This argument is not novel, but it has rarely been taken seriously because it is widely held that the argument has serious flaws. My task in this paper is to argue that these supposed flaws are merely apparent; the Moral Closure Argument is much more potent than it might seem.

1. THE MORAL CLOSURE ARGUMENT

Let us introduce a few of the concepts that will feature prominently in the discussion to come.

Closure: If S knows that P , and P entails Q , and S believes that Q on the basis of competently deducing Q from P , while retaining knowledge of P throughout his reasoning, then S knows that Q .¹

A closure argument is a kind of skeptical hypothesis argument that relies on Clo-

¹ There are other ways to formulate Closure, but this is the most widely accepted version of the principle. For discussion, see Hawthorne, “The Case for Closure.” I assume that Closure is true.

sure (or an instance of the Closure schema) as a premise. To take one famous example: if I know that I have hands, then, by Closure, I would be in a position to know that I am not handless and, therefore, in a position to know that I am not a handless brain in a vat (BIV), or a handless dupe of an evil demon. But I am not in a position to know that I am not being deceived by an evil demon—for all I know, I could be deceived in this way. So, by *modus tollens*, I do not know I have hands. Call this the External World Closure Argument.

The second clause of Closure is “*P* entails *Q*.” Accordingly, in the context of a closure argument, for knowledge of the commonsense proposition to yield knowledge of the falsity of the skeptical hypothesis the skeptical hypothesis must be a scenario where the proposition that is the subject of the skeptical challenge is false. Thus, for a skeptical hypothesis to feature in a closure argument, it must be the case that the skeptical hypothesis is inconsistent with the contested proposition. It follows that skeptical hypotheses are only skeptical *relative to* some proposition or another. The hypothesis that I am a recently envatted brain is a skeptical hypothesis relative to my beliefs about the external world (at the present time), but is not a skeptical hypothesis relative to my beliefs about the past. Conversely, the hypothesis that the world sprung into existence five minutes ago is a skeptical hypothesis relative to my beliefs about the past, but not relative to my beliefs about the present external world.

But skeptical hypotheses do more than stipulate that the proposition in question is false. Skeptical hypotheses also provide an *explanation of our experiences* that is consistent with the falsity of the contested subject matter. The evil demon hypothesis and the BIV hypothesis both provide explanations of my experiences (particularly, my sensory experiences) that are consistent with my beliefs about the external world being false; the five-minute-old-world hypothesis is an explanation of my experiences (particularly, my memory experiences) that is consistent with all my beliefs about the past being false. Let us call this second part of a skeptical hypothesis argument an experience generator.

Some epistemologists, like Stroud and Pryor, have argued that the first condition on a skeptical hypothesis, the falsity condition, is dispensable.² For these epistemologists, a hypothesis can be a skeptical hypothesis even if it is consistent with the truth of the contested proposition. Accordingly, the only thing that is needed for a skeptical hypothesis is the experience generator. Call a skeptical hypothesis that consists in only an experience generator a compatible skeptical hypothesis, because such a skeptical hypothesis is compatible with the truth of the contested proposition. An incompatible skeptical hypothesis is one that is incompatible with the truth of the contested proposition. Only incompatible

2 Stroud, *The Significance of Philosophical Scepticism*; Pryor, “The Skeptic and the Dogmatist.”

skeptical hypotheses can feature in closure arguments, because only incompatible skeptical hypotheses present scenarios whose falsity is entailed by our commonsense beliefs. Because this paper is concerned with closure arguments, we will only consider incompatible skeptical hypotheses.

The distinction between the falsity stipulation and the experience generator of an incompatible skeptical hypotheses and the corresponding distinction between incompatible and compatible skeptical hypotheses will be important to many of the arguments to come. Bear them in mind.

1.1. *The Moral Closure Argument Explained*

The argument that I will defend here—the Moral Closure Argument—mimics the structure of the External World Closure Argument. But instead of showing that we have no external world knowledge, the argument instead shows that we have no moral knowledge.³ And instead of the BIV or evil demon hypotheses, the Moral Closure Argument selects a different skeptical hypothesis.

Sinnott-Armstrong, in his discussion of the Moral Closure Argument, holds that “moral nihilism” is a skeptical hypothesis relative to our moral beliefs.⁴ But that cannot be entirely correct. Moral nihilism—or, more precisely, moral error theory—is, by itself, not a skeptical hypothesis, anymore than the hypothesis “You don’t have hands” is a skeptical hypothesis. Error theory must be part of any *incompatible* skeptical hypothesis relative to our moral beliefs because moral error theory is the only view in metaethics according to which our first-order moral beliefs are *false*.⁵ But we need to supplement error theory with an experience generator.

Fortunately, we already have an explanation of our experiences that is consistent with the falsity of our moral beliefs. Substantial effort has been dedicated—by anti-realists of all stripes, together with empirical psychologists—to showing that our moral intuitions and beliefs can be proximately explained by a myriad of cultural and psychological factors and ultimately explained by evolutionary biology. This evolutionary/cultural/psychological (in short, ECP) story is an experience generator. So our skeptical hypothesis includes both moral error the-

3 Note: *moral* knowledge, not *normative* knowledge. The subject here is moral skepticism, not a broader normative skepticism.

4 Sinnott-Armstrong, *Moral Skepticisms*.

5 Other anti-realist positions in metaethics avoid saying that our moral beliefs are false. Constructivists and relativists hold that our moral beliefs are true—they are just made true by our contingent attitudes or social circumstances. Expressivists either hold that our moral judgments are not truth-valuable, or else that our moral judgments are true in a deflationary sense.

ory and the ECP story. Call this conjunction of error theory and the ECP story Error-ECP. Thus:

Moral Closure Argument

1. I do not know that Error-ECP is false.
2. If I know that killing is wrong, then I know that Error-ECP is false.
3. Therefore, I do not know that killing is wrong.

The argument generalizes at least to any agents like us in situations like ours. (Perhaps God can know that Error-ECP is false, but we cannot.) A general skepticism about morality follows. Note that the Moral Closure Argument is not an argument for moral error theory, but instead an argument against the existence of moral knowledge. Error theory is relevant to this argument only because error theory is an essential part of any incompatible skeptical hypothesis relative to positive moral propositions.

The first premise of any closure argument is motivated by the idea of *evidential underdetermination*. When someone knows that *P*, there is always a way in which they know that *P*. Thus, when someone claims “I know that *P* is false,” a wholly appropriate way to challenge this claim is to ask “How do you know that?” At this point, the purported knower must appeal to some *evidence* that indicates that *P* is false—evidence that would *rule out* the possibility that *P*. But for a skeptical hypothesis, no such evidence is available. Error-ECP entails that our experiences are exactly the same as they are in the actual world. Because of this, none of our experiences gives us any reason to think that Error-ECP is false. And if none of our experiences give us any reason to think that Error-ECP is false, we are not justified in thinking that Error-ECP is false. So we do not know that Error-ECP is false.

That is the logic of the External World Closure Argument as well. The BIV and evil demon hypotheses are hypotheses that entail that your experiences are all exactly the same as they are in the actual world—this is the role of the experience generator. Accordingly, there is (seemingly) no basis on which one might rule out these skeptical hypotheses. So we do not know they are false. The same logic applies to both our inability to rule out Error-ECP and our inability to rule out handless-BIV.

The second premise of the Moral Closure Argument follows from Closure, provided that killing’s being wrong entails that error theory is false and that the subject has competently deduced the falsity of error theory from the wrongness of killing. These two provisos will, generally, be satisfied. Killing’s being wrong does entail that error theory is false. And while some subjects may not recognize the relation between substantive first-order moral claims like “killing is wrong”

and moral error theory, this argument is addressed to those who are able to make the relevant deduction (e.g., academic philosophers). And avoiding the skeptical conclusion by claiming ignorance of basic logic seems a desperate gambit. Thus, if we are to resist the second premise of the Moral Closure Argument, we must reject Closure. But Closure is hard to reject—if P entails Q , and you know that P is true, and recognize the entailment relation that holds between P and Q , and come to believe that Q is true because it is entailed by P , why would you *not* know Q ? Thus: if we knew that killing is wrong, we would be in a position to know that Error-ECP is false.

The Moral Closure Argument highlights the key question that lies at the heart of all skeptical challenges: *How do you know you are not wrong?* The error theorist's position is dismissed (Killing is not wrong? That cannot be right!) more often than it is argued against. Error theorists are entitled to feel a little frustration.⁶ The Moral Closure Argument turns that frustration into a productive use: a demand that error theory be refuted, lest one cede the ground to the skeptic.

In short: we need to have evidence that Error-ECP is false if we are to be justified in believing that any moral claims are true. But we have no evidence that serves to rule out Error-ECP, since that hypothesis provides an excellent account of all our experiences. That is why we have no moral knowledge.

1.2. A Defensive Plan

This marks the end of my positive argument for moral skepticism. My arguments here were brief; closure arguments are well understood. There is no need to belabor the familiar.

What comes next is defense against objections. The objections that I will consider fall into two broad camps. In sections 2–6, I will examine objections that purport to show that the Moral Closure Argument fails for the same reason that the External World Closure Argument fails (whatever that reason might be). In response, I will show that the most popular strategies for responding to the External World Closure Argument systematically fail if we attempt to apply them to the Moral Closure Argument. Despite the similarities between the two arguments, the Moral Closure Argument is a much stronger argument than the External World Closure Argument *because it cannot be answered in the same way*. Of course, the External World Closure Argument has been answered in many different ways, and I cannot survey every possible way to adapt a response to the External World Closure Argument as a response to the Moral Closure Argument. So I do not pretend that my arguments here are conclusive. But by examining a wide range of different, promising anti-skeptical strategies and showing the ways

6 Joyce, "Arguments from Moral Disagreement to Moral Skepticism."

in which they systematically fail to prove useful in answering the Moral Closure Argument, I will show that the Moral Closure Argument is a much more potent skeptical challenge than is typically assumed. It is an argument that anti-skeptics must take seriously.

This discussion will raise to salience a second class of methodological objections, regarding (a) the extent to which the Moral Closure Argument generalizes and (b) whether closure arguments raise independently interesting skeptical concerns. We will consider these objections in sections 7 and 8.

2. THE SPECIAL CASE OBJECTION

The Moral Closure Argument has been considered before, but it has rarely been taken seriously. Sinnott-Armstrong discusses the Moral Closure Argument at some length, but ultimately sets it aside in favor of a discussion of a Regress Argument for moral skepticism.⁷ (According to the Regress Argument, our reasons for our moral beliefs must themselves be justified by prior moral beliefs, which themselves must be justified by prior moral beliefs, and so on *ad infinitum*.) Sinnott-Armstrong holds that the Moral Closure Argument and the Regress Argument are mutually supporting, but he thinks that the Regress Argument is a more illuminating skeptical challenge than the Moral Closure Argument.⁸ And Sinnott-Armstrong is a skeptic; he is *sympathetic* to the Moral Closure Argument. Anti-skeptics tend to be more bluntly dismissive of the Moral Closure Argument. For instance, Michael Huemer writes:

We should not consider it a fair move . . . for someone arguing against ethical intuitionism to deploy general skeptical arguments. . . . Thus, if some particular argument against intuitionism can be shown to be merely a special case of a more general argument impugning our knowledge of those sorts of things, then I may set that argument aside as not relevant to the current discussion.⁹

Call this the Special Case Objection. According to the Special Case Objection, the Moral Closure Argument is defective because it is just a special case of a

7 Sinnott-Armstrong, *Moral Skepticisms*, chs. 3–4.

8 Sinnott-Armstrong's defense of the regress assumes that it is a flaw in any epistemic position if that position begs the question against the skeptic. I hold, along with most contemporary epistemologists, that question-begging anti-skeptical arguments may be sound (see section 4 below). So Sinnott-Armstrong's version of the Regress Argument does not convince. I think Sinnott-Armstrong's Regress Argument can be repaired, but exploring how to do this falls outside the scope of this paper.

9 Huemer, *Ethical Intuitionism*, 12.

more general closure argument schema. Closure arguments are bad; one of the major lessons of twentieth-century epistemology is that the External World Closure Argument is unsound. Accordingly, we have every reason to be suspicious of the Moral Closure Argument; an argument based on a bad argument is also probably bad. This point has been advanced by many others.¹⁰

2.1. Response 1: Validity and Soundness

There are two basic problems with the Special Case Objection. The first is that it reads too much into structural similarities between distinct arguments. If two arguments are structural analogues, both arguments will be *equally valid*; if the premises are related to the conclusion in the same way in both arguments, then, if the premises do not entail the conclusion in the first argument, they will not entail the conclusion in the second argument either. But the logical structure of closure arguments is just a *modus tollens*, and that is a valid argument structure. So the structural similarities between closure arguments give us no grounds for rejecting all such arguments.

If all closure arguments are *unsound*, then, it must be because these arguments have false premises. So the Special Case Objection would be an apt challenge if the Moral Closure Argument and the External World Closure Argument shared premises. But they do not. The first premise of the Moral Closure Argument concerns *Error-ECP*, not BIVs or evil demons. A closure argument against moral knowledge that began with the premise “I know I’m not being deceived by a demon” would be just as suspect as the External World Closure Argument.¹¹ If this is supposed to be the point behind the Special Case Objection, then it is well taken. But in that case, the Special Case Objection does not apply to the Moral Closure Argument. So there is no general reason to think that the Moral Closure Argument will be unsound if the External World Closure Argument is unsound.

2.2. Response 2: Sound Closure Arguments

The second problem with the Special Case Objection is that it implies that *all* closure arguments are unsound, in light of the fact that the External World Closure Argument is unsound. But this is wrong; there are sound closure arguments.

Consider the claim that there are witches. For much of human history, it seemed to many people as though there were witches. But we have been able to

10 See, e.g., Shafer-Landau, *Moral Realism*, 239–40; Star, “Moral Knowledge, Epistemic Externalism, and Intuitionism;” Enoch, *Taking Morality Seriously*, 157–58; McCann, “Conative Intuitionism;” Kulp, “Moral Facts and the Centrality of Intuitions;” Vavova, “Debunking Evolutionary Debunking;” Rosen, “What Is Normative Necessity?”

11 May, “Skeptical Hypotheses and Moral Skepticism.”

explain away all of this seeming evidence. Religious hysteria, social pressure to condemn others as witches, and superstitions all contributed (*inter alia*) to its seeming that there are witches. The claim of the previous sentence is an experience generator. Conjoin that to the negation of the claim that witches exist, and we have an incompatible skeptical hypothesis relative to the claim that witches exist. Call that the No Witch Skeptical Hypothesis. Now consider the following closure argument:

Witch Closure Argument

1. Reverend Parris does not know that the No Witch Skeptical Hypothesis is false.
2. If Reverend Parris knows that Tituba is a witch, then Parris knows that the No Witch Skeptical Hypothesis is false.
3. Therefore, Parris does not know that Tituba is a witch.

This argument is sound. It is also (potentially) compelling. The Witch Closure Argument directly presents Parris with the possibility of error. It makes salient the key question that lies at the heart of all skeptical challenges: *How do you know you are not wrong?* This is a question that Parris should ask himself since, indeed, he is wrong; there are no witches. And by supplementing this possibility of error with a cogent experience generator, it becomes clear that Parris really has no grounds to rule out the No Witch Skeptical Hypothesis. If Parris were to claim that the mere possibility of witchless worlds does nothing to cast doubt on his belief that there really are witches (and, besides, all closure arguments are bad arguments!), he would be missing the point. The Witch Closure Argument is exactly the sort of thing that should make Parris doubt the existence of witches.

We are more familiar with unsound closure arguments than we are with sound closure arguments, because the obviously sound closure arguments, like the Witch Closure Argument, usually are not philosophically interesting. This can bias us to think that all closure arguments contain some deep flaw. But that cannot be right. There are things we should be skeptical about, after all, and closure arguments can help motivate an appropriate skepticism by presenting misguided anti-skeptics with the possibility of error in a compelling way. This is what the Moral Closure Argument does.

3. RELEVANT ALTERNATIVES

If there is no *general* reason to think that the Moral Closure Argument is unsound in virtue of being a closure argument, perhaps there is some *particular* reason to think this. That is, perhaps the flaw that the External World Closure Argument

manifests—whatever that might be—is also present in the Moral Closure Argument. In the next four sections, I will look at four ways in which this suggestion might be substantiated. The four avenues of response that I will consider here are four of the most popular ways of responding to the External World Closure Argument. Thus, while I cannot evaluate every possible response to the External World Closure Argument, the failures of these four responses will show that the Moral Closure Argument is not so easily answered.

3.1. *Objection: Error-ECP Is Irrelevant*

The first particular objection we will examine comes from the Relevant Alternatives framework. According to the Relevant Alternatives framework, a subject knows that P only if that subject can rule out all of the *relevant* alternatives to P . But not every alternative to P is relevant. If a certain proposition, Q , is an *irrelevant* alternative to P , then a subject's inability to rule out Q is no bar to knowing P .

This Relevant Alternatives framework forms the core of a popular response to skeptical hypothesis arguments.¹² All (incompatible) skeptical hypotheses relative to P are *alternatives* to P . But if a skeptical hypothesis is not a *relevant* alternative, then it matters little if we are unable to rule it out. Anti-skeptics can apply this thought to the Moral Closure Argument. If Error-ECP is an *irrelevant alternative*, it need not be ruled out, and thus our inability to rule it out is no bar to our possessing moral knowledge.

3.2. *Response: The Modern Scientific Worldview Must Be Relevant*

Schematically, this suggestion is sensible, and familiar. In Sinnott-Armstrong's discussion of the Moral Closure Argument, Sinnott-Armstrong accepts the Relevant Alternatives framework and, accordingly, structures his discussion around the question of whether Error-ECP ("moral nihilism") is relevant. This depends on what it takes for an alternative to be "relevant" at all. Thus, Sinnott-Armstrong looks hard for an acceptable account of relevance. He cannot find one. Accordingly, he adopts a higher-order, "Pyrrhonian" skepticism: we do not know whether we have moral knowledge because we do not know whether Error-ECP is relevant.

I will not defend an account of relevance here because I am agnostic about the Relevant Alternatives framework. But I do not need an account of relevance; if the Relevant Alternatives framework is correct, then no matter how we understand the notion of a "relevant" alternative, Error-ECP is relevant.

Consider that, for any proposition, P , there is really only one claim that is an alternative to P : $\sim P$. When Relevant Alternatives theorists talk about *different*

12 Dretske, "Epistemic Operators."

alternatives to P , they are not talking about propositions other than $\sim P$. They are talking about different ways for P to be false—or, more precisely, different stories about why it might seem that P is true, even though it is false. Thus, if we accept the Relevant Alternatives framework, what we are evaluating for relevance is the experience generators of different skeptical hypotheses. The irrelevant alternatives will be alternatives that appeal to experience generators that are *outlandish*, to use Sinnott-Armstrong's apt term. The idea that you could be a BIV is an outlandish one, indeed.

Thus, if we are to apply the Relevant Alternatives framework to rule out Error-ECP, it must be on the grounds that the experience generator of Error-ECP is outlandish. But the experience generator of Error-ECP is the ECP story, and that story is not an error theorist's wild conjecture. It does not talk about demons or brains in vats. It is, instead, an empirically verified story, and the consensus view among psychologists who study the origins of our moral beliefs.¹³ This scientifically respectable story is not outlandish in any sense of the word. So Error-ECP must be a relevant alternative to our moral beliefs, if anything is.

One might object that Error-ECP *really is* an outlandish view. What makes it outlandish is not its commitment to a scientific explanation of our experiences, but the fact that it is *a version of error theory*. The outlandish feature of Error-ECP is Error, not ECP.

This is an important objection, but it is not based on the Relevant Alternatives framework. The Relevant Alternatives framework provides a way of evaluating the *experience generators* of skeptical hypotheses. Responses to skepticism that proceed by arguing that it is absurd that our commonsense beliefs might be false are *Moorean* responses to skepticism. I turn to those next.

4. MOOREANISM

Mooreanism is the view that we can know that skeptical hypotheses are false in virtue of the fact that we have basic knowledge that certain of our first-order beliefs are true. A Moorean about the external world says that she can know she is not a BIV because she has basic knowledge that she has hands. A Moorean about morality, by like token, is someone who says that she can know that Error-ECP is false because she has basic knowledge that killing is wrong (for example).

The most common objection to Moorean responses is that they beg the question. In making a claim to basic knowledge that P , Moorean responses to skepticism begin by assuming the truth of P ; this is blatantly question begging,

¹³ See Joyce, *The Evolution of Morality*, chs. 1–2.

and that seems like a bad thing.¹⁴ Sinnott-Armstrong structures his discussion of moral skepticism around the question of whether any arguments can be advanced that do not beg the question against a moral nihilist.¹⁵

But it is not obvious that begging the question is bad. Question-begging arguments may be *sound*, after all. Further, the premises of sound question-begging arguments might be premises *that I know to be true*, even if my interlocutor does not believe them. So what, then, is the problem with making a question-begging argument, provided that it is valid and that I know the premises to be true? The only problem seems to be that it cannot convince dedicated skeptics to abandon their skepticism.¹⁶ But, of course, it is impossible to convince people who are dead set against being convinced. So the only flaw in question-begging arguments is that they do not accomplish an impossible task; that does not look like much of a flaw.

This does not mean that the Moorean response is in the clear. Question-begging arguments are fine, provided that the argument is valid and that I know the premises to be true. But how can the anti-skeptic substantiate her claim to knowing the commonsense proposition in question, in light of the skeptical argument? That is the burden the Moorean must shoulder.

One way to substantiate this Moorean strategy is to offer a *particularist* response to the Problem of the Criterion. The Problem of the Criterion is an ancient methodological puzzle within epistemology—how do we know the extent of our knowledge unless we first have an idea of what knowledge is? And how do we know what knowledge is unless we first have an idea of the kinds of things we are talking about when we are theorizing about knowledge? It seems like we cannot answer either question until we have first answered the other. The particularist grasps the first horn of this dilemma and holds that our knowledge of particular matters of fact is absolutely methodologically prior to our acceptance of any epistemic principles. So particularism provides us with one way to be Mooreans—by holding that claims like “I know that killing is wrong” or “I know I have hands” are methodological fixed points in epistemic theorizing.

I have nothing to say in response to such a Moorean particularist—how could I? Particularists are methodologically committed to not abandoning their claims to knowledge, and as we just observed, there is no convincing people who are dead set against being convinced. Yet it is this very inflexibility of particularism that makes it such an unattractive solution to the Problem of the Criterion. The best response to the Problem of the Criterion seems to me to be a response

14 Joyce, “Arguments from Moral Disagreement to Moral Skepticism.”

15 See note 8 above.

16 Pryor, “What’s Wrong with Moore’s Argument?”

that denies that either particular claims or general principles have absolute priority over the other. The best epistemic theory is to be found in a state of reflective equilibrium between particulars and principles. This means that it is incumbent upon Mooreans to do more than insist on their commonsense claims. They should give some account of why the claims that they are insisting on are the right claims to insist on.

The simplest way to do this was suggested by Moore himself. According to Moore, we should view skeptical arguments as presenting a kind of *aporia*, where the truth of the premises and the negation of the conclusion all seem plausible, yet one of those claims must be false. And, in such a situation, the thing to do is to reject the claim in which we have the lowest degree of confidence. Thus, the skeptic can be dismissed on the grounds that “It seems to me *more* certain that I *do* know that this is a pencil and that you are conscious, than that a single one of these four assumptions [of the skeptical argument] is true, let alone all four.”¹⁷ “Confidence may serve as defeasible *indirect* evidence of the truth of a claim.”¹⁸

But it is implausible that *mere* confidence serves as evidence of any kind in favor of a claim. People can be (and often are) irrationally confident, and irrational confidence in a claim is no evidence for the truth of that claim. How confident an agent ought to be in a claim is a function of what evidence the agent has for that claim, not the other way around. If we were to present Reverend Parris with a modern scientific explanation of all the things that he takes to be evidence that Tituba is a witch and ask him, in light of the availability of this scientific explanation, what reason he has to think that Tituba is a witch, it will not do for Parris to respond by avowing his unwavering confidence. We need evidence of a kind to substantiate the Moorean position, but mere confidence will not do.

4.1. Moorean Dogmatism

The most well-developed and influential version of Mooreanism—dogmatism—addresses this problem. Dogmatism was developed by Jim Pryor and Michael Huemer as a response to the External World Closure Argument, and in later work, Huemer expanded his account to address the Moral Closure Argument as well.¹⁹ This makes Moorean dogmatism a strong candidate for the best response to the Moral Closure Argument. The cornerstone of the dogmatist account is a principle called

17 Moore, quoted in McPherson, “Moorean Arguments and Moral Revisionism,” 4.

18 McPherson, “Moorean Arguments and Moral Revisionism,” 5.

19 Pryor, “The Skeptic and the Dogmatist” and “What’s Wrong with Moore’s Argument?”; Huemer, *Skepticism and the Veil of Perception and Ethical Intuitionism*, ch. 5.

Phenomenal Conservatism (PC): If it seems to S as if P, then S has at least *prima facie* justification for believing that P.²⁰

PC contains two important notions. First is the notion of P's *seeming* true to S. *Seemings* are mental states that have propositional content, thereby presenting the world as being a certain way and (according to PC) justifying our belief that the world is that way. Sensations are *perceptual* seeming states. Memory and introspection are also seemings, although of different kinds. And there is a fourth kind of seeming, an *intellectual* seeming state, where some proposition just seems true on reflection. These intellectual seeming states are *intuitions*. And *moral intuitions* are just intellectual seeming states whose content is a moral proposition.²¹

According to PC, our sensations (perceptual seeming states) present the external world as being a certain way; consequently, we are *prima facie* justified in believing the world is that way. In similar fashion, our moral intuitions present certain moral propositions as being true; consequently, we are *prima facie* justified in believing those propositions. Thus, while our experiences might be the same in both a skeptical scenario and non-skeptical scenario, the *content* of our experiences uniquely supports the non-skeptical hypothesis, by PC.

To say that one is *prima facie* justified in believing P is only to say that the agent has a defeasible reason to believe P. But *undefeated prima facie* justification is sufficient for all-things-considered justification. Thus, having a seeming state that P is different from merely being confident that P, but an undefeated seeming that P *makes it rational to be confident* that P. If you are *prima facie* justified in believing that P, and there are no defeaters for your belief, then you are all-things-considered justified in believing that P. In this way, we may come to know that the commonsense hypothesis is true and so, by Closure, may come to know that the skeptical hypothesis is false.

4.2. Response: Undermining Prima Facie Justification

I will grant, *arguendo*, that PC is true and that moral intuitions are seeming states with moral contents. From this, it follows that we have some *prima facie* justified moral beliefs. Both of those premises are controversial, but I will accept them anyway because Huemer makes no case that our moral beliefs are undefeated. And this is how Huemer's argument fails: there are defeaters for our moral beliefs that are not present for our external world beliefs.

So, for Huemer and Pryor, what does it take for *prima facie* justification to

20 This formulation is Huemer's; Pryor accepts a restricted form of this principle.

21 Huemer, *Ethical Intuitionism*, ch. 5.

be defeated? Before we answer that question, a reminder: in section 1, I distinguished between a compatible and an incompatible skeptical hypothesis, and showed that only incompatible skeptical hypotheses can feature in closure arguments. But Huemer and Pryor dislike closure arguments, and so interest themselves in *compatible* skeptical hypotheses. When Pryor and Huemer talk about a skeptical hypothesis, they are talking about what I have been calling an *experience generator*: to wit, an explanation of your experiences that is consistent with the falsity of the proposition in question. Keeping that in mind, here is Pryor on what it would take to defeat *prima facie* justification:

This *prima facie* justification can be undermined or threatened if you gain positive empirical evidence that *you really are in a skeptical scenario*. (For instance, if a ticker tape appears at the bottom of your visual field with the words “You are a brain in a vat . . .”) If you acquire evidence of that sort, then you’d have to find some non-question-begging way of ruling the skeptical hypothesis out, before you’d be *all things considered justified* in believing that things are as your experiences present them. In the standard case, though, when the *prima facie* justification you get from your experiences is not defeated or undermined, then it counts as all things considered justification, without your having to do this.²²

Huemer agrees, saying: “Does the brain-in-a-vat hypothesis, itself, constitute such a defeater? It would, if I had some reason for accepting it—that is, if I had significant evidence that I am a brain in a vat. But I have no such evidence, nor have the world’s skeptics proven interested in trying to provide that sort of evidence.”²³

Pryor and Huemer both provide one example of what *would* defeat *prima facie* justification: evidence in favor of a (compatible) skeptical hypothesis, or what I have been calling an experience generator. And this seems right. If you learn that you are dreaming, that undermines your perceptual justification for thinking that you have hands, even though the dreaming hypothesis is compatible with your having hands. So Pryor and Huemer’s Moorean account rests on the claim that we have no evidence in favor of an experience generator (like the BIV hypothesis) that would undermine our justification for thinking that we have hands. A ticker running across the bottom of our visual field telling us that we are brains in vats (indeed, *any* ticker running across the bottom of our visual field) would be hard to account for on the hypothesis that our senses put us

22 Pryor, “The Skeptic and the Dogmatist,” 537–38. Pryor reiterates this suggestion in “What’s Wrong with Moore’s Argument?”

23 Huemer, *Skepticism and the Veil of Perception*, 183.

in contact with an external world. Accordingly, experiences like that would be good (if not conclusive) evidence that some skeptical hypothesis is true. Fortunately, we have no evidence like that.

But we cannot say the same thing about the moral case. Our *prima facie* justified moral beliefs are defeated if we have “positive empirical evidence” in favor of an experience generator of a skeptical hypothesis relative to our moral beliefs. And we do have evidence like this! The experience generator for the Moral Skeptical Hypothesis is the ECP story, and that story is supported by copious empirical research. Thus, there is a defeater for the *prima facie* justification provided by our moral intuitions.

I am not arguing here that evidence for an experience generator is *generally* a defeater for *any* of our beliefs. (I do think that is true, but I argue for it elsewhere.)²⁴ Rather, the point is that Pryor/Huemer-style dogmatism cannot be used to substantiate a Moorean response to the Moral Closure Argument. Pryor’s and Huemer’s version of Mooreanism rests on a principle of *prima facie* justification that contains a proviso which states conditions under which justification will be defeated. As Pryor and Huemer both correctly point out, that proviso is unsatisfied in the case of our external world knowledge: we have no reason to think that any experience generator relative to our beliefs about the external world is true. However, we *do* have reason to think that the experience generator of Error-ECP is true. So the proviso is satisfied in the case of the Moral Closure Argument. The dogmatist response fails.

5. INFERENCE TO THE BEST EXPLANATION

Another popular response to the External World Closure Argument appeals to Inference to the Best Explanation (IBE). According to this response, the hypothesis that there is an external world that is largely the way that we perceive it to be is the best explanation of our experiences (particularly, our sensory experiences).²⁵ Thus, by IBE, we can know that the anti-skeptical hypothesis is true and that the skeptical hypothesis is false.

A similar argument can be given against the Moral Closure Argument. According to some moral realists (most notably the “Cornell realists”), moral facts can explain other kinds of facts—including, potentially, facts about our expe-

²⁴ Lutz, “What Makes Evolution a Defeater?”

²⁵ Cf. Russell, *The Problems of Philosophy*, ch. 2; Vogel, “Cartesian Skepticism and Inference to the Best Explanation.” This is, incidentally, the response to the External World Closure Argument to which I am most sympathetic.

riences. And if moral facts can explain our experiences, it is possible for moral facts to constitute the *best explanation* of our experiences.

This position has been developed by Majors in response to Sinnott-Armstrong's Regress Argument.²⁶ Sinnott-Armstrong argues that moral facts do not feature in the best explanation of anything: our best explanations will appeal only to *nonmoral base properties*, not to moral properties.²⁷ Majors responds by saying that moral properties are a kind of higher-order, supervenient natural property.²⁸ These higher-order properties are individuated by their distinct explanatory profile. If we want to know why a revolution occurred, we might explain this occurrence by giving a detailed description of the complete political and social circumstances just prior to the revolution; but the *best explanation* of that revolution is the higher-order property that these political and social circumstances constitute *injustice*. This higher-order explanation is the most accurate, informative explanation of the revolution, because it is the explanation that captures the appropriate degree of generality—revolutions are far more likely to occur in political and social conditions that manifest injustice than in those that do not.²⁹

Thus, Majors argues that because higher-order properties have robust explanatory power at the right level of generality, IBE will give us a reason to believe in these higher-order moral properties and reject Sinnott-Armstrong's account of explanations in terms of nonmoral base properties.³⁰ By IBE, we can know that killing is wrong, and thus, by Closure, come to know that Error-ECP is false. The Moral Closure Argument fails.

5.1. Response: Higher-Order Nonmoral Properties

This debate between Sinnott-Armstrong and Majors is confused. Sinnott-Armstrong's position is that *nonmoral base properties* can do all the explanatory work; Majors's counter is that *higher-order moral properties* are needed to do explanatory work. But even if Majors is correct to say that we need to refer to higher-order properties to capture the right level of generality in our explanations—and I do find this position compelling—then this only shows that we must have higher-order properties in our explanations. It does not show that those higher-order properties must be *moral* properties.

Error-ECP is not a view that denies the existence of higher-order properties

26 Majors, "Moral Explanation."

27 Following Harman, *The Nature of Morality*.

28 Following Sturgeon, "Moral Explanations."

29 Railton, "Moral Realism."

30 See also Oddie, *Value, Reality, and Desire*.

with explanatory power. Many of the properties that feature in the ECP story are higher-order properties of the kind that Majors is interested in. So the relevant comparison between explanations, for our purposes, is a comparison between a view on which there are higher-order moral properties with explanatory power, and a view on which there are higher-order properties with explanatory power *that are not moral*. It is easy to miss this distinction because many of the concepts that we use to pick out such higher-order properties are “thick” normative concepts—INJUSTICE is a perfect example. Thick concepts like INJUSTICE have both normative and descriptive significance. But these normative and descriptive aspects of the concepts are conceptually separable. Let us call the descriptive aspect of the concept of injustice DESCRIPTIVE-INJUSTICE, and the property that is picked out by this concept descriptive-injustice. According to Error-ECP, there is a nonmoral property of descriptive-injustice that (*inter alia*) explains why political revolutions tend to happen in societies that have this property, and tend not to happen in societies that lack it.

For comparison, consider ethnic slurs. Some bigots have the concept of a MICK, but we can pick out the descriptive aspect of this concept—PERSON FROM IRELAND—which refers to the nonnormative property of being from Ireland. The property of being from Ireland has explanatory power (e.g., it explains the accent). But the fact that this property has explanatory power gives us no reason to believe in the normatively valenced property of being a rotten *mick*. Similarly, some social organizations—the descriptively-unjust ones—might tend toward political revolution. But the explanatory power of descriptive-injustice gives us no reason to believe in a normatively valenced property of *injustice*.

If the anti-skeptic seeks to use IBE to rule out Error-ECP, that means that we need to ask whether a normatively valenced property of injustice explains our observations of patterns in political revolutions better than explanations couched entirely in terms of descriptive-injustice. Thus, the only way to use IBE to argue against Error-ECP is to argue not just that there are higher-order properties with robust causal profiles that can explain events like political revolutions, but *also* that those properties have a moral valence, and that this moral valence adds to their explanatory power. But we have no reason to accept these further claims. The fact that we should believe in *higher-order* properties is no evidence that we should believe in *moral* properties.

5.2. *The Epistemic Quality of Error-ECP*

The arguments of the last three sections have a common theme. In assessing whether an alternative is “relevant” or not, we need to look to the experience generator of the skeptical hypothesis in question. Because the experience gener-

ator of Error-ECP is a scientifically confirmed account, rather than an outlandish conjecture about demons or brains in vats, the Relevant Alternatives response fails. Moorean accounts of justification only succeed in providing all-things-considered justification if we have no evidence in favor of an experience generator of a skeptical hypothesis. But when the skeptical hypothesis in question is Error-ECP (rather than handleless-BIV), we do have evidence for the experience generator. And IBE is no help in answering the Moral Closure Argument because the experience generator of the Moral Closure Argument is, as a scientifically confirmed theory, a *product* of reasoning in accordance with IBE.

The point is not that evidence in favor of the ECP story is a reason to believe that there are no moral facts. As I argued in section 1, the intuitive force of the Moral Closure Argument rests on the fact that the anti-skeptic has no good answer to the question “How do you know you are not wrong?” Yet the most prominent answers to the External World Closure Argument all, in one way or another, indicate how strange and implausible it is to suggest that we are dupes of an evil demon or brains in vats. But to apply these same counterarguments to the Moral Closure Argument, the anti-skeptic will have to argue that it is strange and implausible to suggest that our beliefs are the product of the ECP story. But that is not implausible at all.

This is similar to a point raised by McPherson.³¹ McPherson (who is neither an error theorist nor a skeptic) argues that Moorean responses cannot rule out moral error theory. Moorean responses only serve to rule out a skeptical hypothesis if that skeptical hypothesis lacks “generic indicators of epistemic quality.”³² And, McPherson argues, moral error theory does not lack generic indicators of epistemic quality. McPherson’s account of the generic indicators of epistemic quality is different from my own, and he does not appreciate the importance of the distinction between falsity stipulations and experience generators within an incompatible skeptical hypothesis. But his anti-Moorean argument is essentially correct.

At this point, one is liable to have a number of worries about how far this skeptical argument generalizes and how it relates to other kinds of skeptical arguments, particularly evolutionary debunking arguments. I address these concerns in sections 7–8. For now, there is one more important response to the Moral Closure Argument to examine.

31 McPherson, “Moorean Arguments and Moral Revisionism.”

32 McPherson, “Moorean Arguments and Moral Revisionism,” 6.

6. EXTERNALISM

We are currently examining popular ways of responding to the External World Closure Argument, and seeing whether they can be adapted to serve as responses to the Moral Closure Argument. In the previous two sections, we looked at *internalist* responses to skepticism, which ground justification in either the content of, or the best explanation of, our *experiences*. In this section, we will look at *externalist* responses to skepticism, which hold that we are justified in rejecting error theory in virtue of things external to our experiences.

The externalist offers a simple and compelling thought: because a good experience generator will provide an adequate explanation of all of our experiences, we have no internally accessible grounds for ruling out skeptical hypotheses. Yet we clearly do know things; it follows, then, that our beliefs are justified in virtue of facts that are *not* internally accessible to us. If we are being deceived, then our beliefs are not justified. But if we are not being deceived, then our moral beliefs are justified (e.g., in virtue of our reliability). This thought can be applied in the domain of morality just as easily as in other domains. If we have no internally accessible basis for ruling out Error-ECP, but are nonetheless justified in holding moral beliefs (and of course we are justified in holding moral beliefs!), then we must be justified in holding those beliefs on external grounds.³³

6.1. Response: The New Evil Demon Problem

The main problem with externalist responses to skepticism is that they incorrectly predict that we have different evidence depending on whether or not we are being deceived. That commitment is the core of the externalist anti-skeptical project. But it is counterintuitive, as the (rightly) famous New Evil Demon Problem shows.³⁴ If two agents have all the same experiences and background beliefs, those two agents will be justified in believing all the same things, regardless of whether they are forming true beliefs or not. Someone who is being deceived by an evil demon, but does not know it, is just as justified in their beliefs as someone who is not being deceived at all. Provided that the demon's illusion is sufficiently comprehensive, it would be irrational for the deceived individual to do anything other than believe in accordance with the demon's illusion. So while the epistemic externalist claims that we are justified in believing fewer things in cases where we are deceived, that just seems wrong.

Star holds that epistemic internalism lacks the resources to overcome moral

33 Star, "Moral Knowledge, Epistemic Externalism, and Intuitionism."

34 Cohen, "Justification and Truth"; Wedgwood, "Internalism Explained."

skepticism, and thus we must be externalists or else accept moral skepticism.³⁵ Star takes this to be conclusive reason to be an epistemic externalist. I agree that our inability to provide an internal justification of our moral beliefs forces us to choose between either externalism or moral skepticism. However, in light of the New Evil Demon Problem, moral skepticism seems to be the correct conclusion to draw.

6.2. Evidence and Experience

The New Evil Demon Problem presents us with an intuitive datum: that our evidence, and thus our justification, is the same whether or not we are being deceived (provided the deception is sufficiently comprehensive). Externalists have responded to this challenge in two different ways. According to what we may call the *concessive response*, our beliefs in an external world are *equally rational* whether or not we are deceived, but only in cases where we are not deceived are our beliefs *fully justified*.³⁶ Concessive externalists are correct to say that there are epistemic goods (like, perhaps, reliability) that are not internally accessible to us. But in conceding that it is not rational to believe that one is being deceived in cases of sufficiently comprehensive deception, a concessive externalist is conceding that our experiences are not sufficient to rule out the skeptical hypothesis. That is enough to get a closure argument off the ground.

A second externalist response, which we may call the *knowledge first response*, comes from Timothy Williamson.³⁷ Williamson holds that the only true epistemic good is knowledge, and that our evidence consists in our knowledge; his slogan is “ $E = K$.” If $E = K$, then, because we have knowledge in the “good case” (where we are not being deceived) and lack knowledge in the “bad case” (where we are being deceived), we have different evidence in those two cases. This provides a potential answer to the Moral Closure Argument with a Moorean flavor: because we know many moral facts, those known moral facts serve as evidence that rules out ERROR-ECP.

Williamson attacks the idea that our evidence consists in our experiences in two ways. First, he presents an argument “on behalf of the skeptic” that purports to prove that our evidence is the same in the good case as in the bad case. He then shows that this argument fails.³⁸ Williamson’s refutation of his own argument is compelling, but it does not address the New Evil Demon Problem. The New Evil Demon Problem is not supposed to *prove* that evidence is experience.

35 Star, “Moral Knowledge, Epistemic Externalism, and Intuitionism.”

36 See Bonjour and Sosa, *Epistemic Justification*, 162–65.

37 Williamson, *Knowledge and Its Limits*.

38 Williamson, *Knowledge and Its Limits*, ch. 8.

It is a burden-shifting argument, intended to show how counterintuitive the externalist account of evidence is. Williamson owes us more.

This comes in the form of a positive argument for $E = K$. However, this argument is weak. One of the premises in Williamson's argument for $E = K$ is that all evidence is propositional. Williamson supports this by arguing that explanatory reasoning is a paradigmatic way by which beliefs can be justified, and only propositions can be the relata of explanations. It makes no sense to demand explanations of *objects*.³⁹ But Williamson is wrong to assume that objects and propositions are the only two things that we might be attempting to explain when we engage in explanatory reasoning. When we engage in explanatory reasoning, we are (often) reasoning about the best explanation of *events*—including events that consist in our having experiences.⁴⁰ For instance, my feeling warm (viz., an event of my having a certain warm experience) is best explained by the room being warm, and thus this experience is evidence that the room is warm. Indeed, *contra* Williamson, it is not clear what would explain an abstract entity like a proposition. For Williamson, propositions are sets of possible worlds; what does it mean for something to explain a set of possible worlds? I have a warm experience if and only if the proposition \langle I have a warm experience \rangle is true. But it is not the proposition that is explained, but the event of my feeling warm.⁴¹

Williamson also argues that we should reject non-propositional evidence because “it is hard to explain how non-propositional evidence contributes to updating probabilities.”⁴² But this, too, fails to convince. First, those who endorse evidence-as-experience have put substantial effort into explaining how non-propositional evidence can contribute to updating probabilities.⁴³ Second, Williamson's own account of how propositions contribute to updating depends on his assumption that propositions are sets of possible worlds; but this is a controversial assumption.⁴⁴ Third, it is not at all clear that evidential support should be analyzed in terms of “updating probabilities.”⁴⁵ And fourth, Williamson's objection threatens to confuse the map with the territory: even if it is a necessary condition on a notion of evidence that we must be able to model it probabilisti-

39 Williamson, *Knowledge and Its Limits*, 194–95.

40 Conee and Feldman, “Replies.”

41 McCain, *Evidentialism and Epistemic Justification*, 14–15.

42 Williamson, “Knowledge First,” 9.

43 Dougherty and Rysiew, “Experience First”; Poston, *Reason and Explanation*.

44 For a number of arguments against the view that propositions are sets of possible worlds, as well as arguments in favor of rival conceptions of propositions, see King, Soames, and Speaks, *New Thinking about Propositions*.

45 Cf. Horgan, “Troubles for Bayesian Formal Epistemology.”

cally, the fact that propositions serve as inputs in a probability calculus does not prove that evidence *consists in* propositions, but only that we should be able to model a correspondence between evidence and propositions.

In sum: the New Evil Demon Problem remains a strong reason to reject externalist accounts of evidence and rationality. Williamson's argument for $E = K$ is weak; it provides no good reason to reject the intuitive idea that our evidence consists in our experiences. Thus, externalism is not a promising way to respond to any closure argument, including the Moral Closure Argument.

7. OVERGENERATING SKEPTICISM

I have argued in the last four sections that the Moral Closure Argument is compelling because the standard responses to the External World Closure Argument do not apply to the Moral Closure Argument, and thus do not rebut moral skepticism. But one might worry that my arguments threaten to prove too much. One of my main points has been that the skeptical hypothesis that features in the Moral Closure Argument has an experience generator that is an empirically confirmed product of the modern scientific worldview. But this line of argument threatens to overgenerate skeptical conclusions along two different lines: it seems to support skepticism about macrophysical objects, and skepticism about the future. In this section, I will show how closure arguments for skepticism about macrophysical objects and about the future can be answered in ways that do not impute the soundness of the Moral Closure Argument.

7.1. Objection: Skepticism about Ordinary Objects

Let us formulate an incompatible skeptical hypothesis relative to the proposition that there are tables, in accordance with the procedure outlined in section 1. First, we stipulate that there are no tables. This, by itself, is not a skeptical hypothesis; we need an experience generator. Unfortunately, we can find one in the modern scientific worldview. The modern scientific worldview tells us that our experiences can be explained in terms of *fundamental particles arranged tablewise*. But it is *logically possible* that there can be fundamental particles without there being tables. So a hypothesis according to which there are fundamental particles but no tables—i.e., eliminativism about ordinary objects—is a skeptical hypothesis relative to the claim that there are tables. We have no way to rule that out. So we do not know that there are tables. Call this the Ordinary Objects Closure Argument.

The fact that we are talking about particles *arranged tablewise* means that we are talking about a property with substantial explanatory power at the right level

of generality to explain our observations of and judgments about tables. Claiming that those particles-arranged-tablewise *are tables* adds nothing to the explanatory power of that hypothesis. So IBE gives us no reason to prefer the Arranged Particle Hypothesis to the hypothesis that there are ordinary objects. This is an application of my response to Majors from section 5. Thus, if my arguments there were successful, then it seems a similar strategy can be used to defend the Ordinary Objects Closure Argument. That seems bad.

7.2. Response: Analyticity

But we do know that eliminativism about ordinary objects is false, because that hypothesis is conceptually incoherent. It is *analytic* that, if there are particles arranged tablewise, then there are tables.⁴⁶ According to Thomasson, the relation between our concept of a TABLE and our concept of PARTICLES ARRANGED TABLEWISE is an analytic relation *par excellence*—the predicate “table” is literally contained in the predicate “particles arranged tablewise.” So it is *conceptually impossible* for there to be particles arranged tablewise without there being tables. Accordingly, any conceptually competent individual can know that the eliminativist hypothesis is false by reflecting on the constituent concepts.⁴⁷ I find this argument compelling.

But this same response is not available to anti-skeptics about morality. Error-ECP is not conceptually incoherent. Our concepts of natural properties do not contain the concepts of moral properties. To conceive of the world as containing particles arranged tablewise is, *ipso facto*, to conceive of the world as containing tables. But to conceive of the world as containing killings is not, *ipso facto*, to conceive of the world as containing immoral actions. This is one important lesson from Moore’s Open Question Argument.⁴⁸ “X is particles arranged tablewise, but is X a table?” is a closed question—to ask it betrays a lack of conceptual competence with “particles arranged tablewise.” But “X is a killing, but is X wrong?” is not a closed question—to ask it betrays no lack of conceptual competence with “killing.”

One might reject Thomasson’s account of our concepts and ordinary objects—and her account is controversial.⁴⁹ A full defense of Thomasson falls outside the scope of this paper. But if Thomasson’s argument fails, skepticism about ordinary objects begins to look much more attractive (to me, at least).⁵⁰ So ei-

46 Thomasson, *Ordinary Objects and Ontology Made Easy*, ch. 3.

47 Thomasson, *Ontology Made Easy*.

48 Moore, *Principia Ethica*; cf. Bedke, “Against Normative Naturalism.”

49 Cf. Button, “Deflationary Metaphysics and Ordinary Language.”

50 See also McPherson, “Moorean Arguments and Moral Revisionism,” sec. 3.

ther Thomasson's argument succeeds, or the Ordinary Objects Closure Argument is sound. Either way, the Moral Closure Argument does not overgenerate skepticism.

7.3. *Objection: Future Outlandishness*

Let us formulate a skeptical hypothesis relative to some claim about the future: say that, in 2050, the Rocky Mountains will spontaneously transform into a giant whale.⁵¹ This entails that my belief that the Rocky Mountains will undergo no such transformation several decades hence is false. That is not, by itself, a skeptical hypothesis relative to my future belief; we need an experience generator. Unfortunately, we can find one in the modern scientific worldview. Take whatever explanation of my experiences that modern science has to offer; it is *logically possible* that this explanation is true and for the Rocky Mountains to transform into a giant whale. That is because this experience generator is cashed out in terms of *past* facts, which is consistent with the claim that the Rockies will become a giant whale *in the future*. So this Giant Whale Skeptical Hypothesis is a skeptical hypothesis relative to a future belief. We have no way to rule that out. So we do not know that the Rockies will fail to transform into a giant whale. Call this the Giant Whale Closure Argument. This argument will generalize to entail a skeptical conclusion about all future facts, because the past explanation of my experiences will always be consistent with the falsity of my future beliefs. If my strategy in defending the Moral Closure Argument has been successful, then it seems a similar strategy can be used to defend the Giant Whale Closure Argument. That seems bad.

7.4. *Response: Induction*

But we can know that the Giant Whale Skeptical Hypothesis is false by inductive reasoning. The Rocky Mountains have not spontaneously transformed into a giant whale at any point in the past; indeed, nothing has ever spontaneously transformed into a giant whale at any point in the past. And we know why this should be the case, as our best physics gives us an excellent understanding of the ways in which matter changes over time. This long history of mountains not transforming into whales, combined with our understanding of the causal processes that explain this history, give us an excellent inductive basis from which we might infer that the Giant Whale Skeptical Hypothesis is false.

But this same response is not available to anti-skeptics about morality. Inductive reasoning works by extending a pattern of observed facts into unobserved instances. To have moral knowledge by induction, we must first have non-in-

51 I owe this amusing hypothesis, and the accompanying objection, to an anonymous referee.

ductive moral knowledge. If my arguments thus far have been successful, that is precisely what we lack.

One might worry that Hume's argument for inductive skepticism will create difficulties for this response. Of course the best explanation of our experiences entails that mountains do not transform into whales, but this is only true *in the past*. But what reason do we have to think that the future will resemble the past? If we think of the Giant Whale Skeptical Hypothesis as an instance of counter-inductive reasoning, then it is question begging to say that we can rule out that hypothesis by inductive reasoning.

This is an important objection, but it can be answered; we can justify induction inductively. Inductive justifications of induction are question begging, but it is no vice for anti-skeptical arguments to beg the question (section 4).⁵²

One might also worry that I am making things too easy for myself by focusing on the Giant Whale Skeptical Hypothesis. It is, perhaps, reasonable to inductively reject *that* skeptical hypothesis, but we might have less ground to rule out skeptical hypotheses about the future that are less outlandish. Thus, my arguments here still may direct us toward skepticism about many other future facts. This charge has merit, but it is not an objection. Predictions are hard, especially about the future. I can know that the sun will rise tomorrow and that mountains will not spontaneously transform into whales. About many other things, it is best to suspend judgment.

8. MAY'S OBJECTION

When the Moral Closure Argument is discussed in the literature on moral epistemology, it is often dismissed with the Special Case Objection. But Jonathan May has provided a subtler critique of the Moral Closure Argument.⁵³ I turn to this last.

May's concern begins from the observation that the first premise of a closure argument—that we do not know the skeptical hypothesis to be false—is always the most contentious. Once we have established the first premise of the closure argument, the conclusion follows pretty easily. This means that the real force of a closure argument comes from the considerations advanced in favor of the

52 Inductive justifications of induction are controversial, of course, but pursuing this controversy in depth will take us too far afield. For defenses of an inductive justification of induction, see Alston, "Epistemic Circularity," and Van Cleve, "Reliability, Justification, and the Problem of Induction." I argue in favor of an inductive justification of induction in other work (Lutz, "Defusing the Counterinduction Parody").

53 May, "Skeptical Hypotheses and Moral Skepticism."

first premise of the argument. May claims that this is a sufficient reason to set the closure argument aside—discussions of Closure and skeptical hypotheses are a distraction from the core concern, which is the motivation behind the first premise. And what is the motivation behind the first premise? That depends on what the first premise of the Moral Closure Argument is. May argues that if the moral skeptic appeals to hypotheses about *brains in vats* or *evil demons* or the like, then the Special Case Objection seems to apply (section 2.1 above). But if the first premise of the Moral Closure Argument is *ERROR-ECP*, then the force behind the Moral Closure Argument comes from the fact that we can give evolutionary explanations of our experiences. And this means that the Moral Closure Argument is nothing more than an evolutionary debunking argument, on which there is already a vast literature. The Moral Closure Argument adds nothing new.

This kind of worry may seem to be particularly pressing against the version of the Moral Closure Argument I have defended here. In sections 3–5, I emphasized that *ERROR-ECP* is a realistic and not-at-all-outlandish hypothesis because it is a hypothesis that is supported by the available evidence, giving it a high degree of epistemic quality. So if my argument here is successful, it seems to be because we have evidence that our moral beliefs are the product of evolution. This, argues May, makes it a kind of evolutionary debunking argument.

8.1. Not Quite Debunking

There are two ways to understand the force behind May's objection. The first is that the evolutionary debunking argument serves as a lemma within the Moral Closure Argument, as it provides the justification for its first premise. Thus, the skeptical force of the Moral Closure Argument would be parasitic on the success of a debunking argument; the closure argument structure adds nothing. This is May's stated concern.

However, the first premise of the Moral Closure Argument is motivated not by a debunking argument but by the idea that we have no evidence against the error theorist's total worldview (error theory plus whatever the best of natural science has to tell us about the origins of our moral beliefs). As Vavova argues, there is a difference between evolutionary debunking arguments, which attempt to show that evolutionary explanations of our moral beliefs *give us a good reason* to think that our moral beliefs are (probably) *false*, and the Moral Closure Argument, which attempts to show that our experiences *give us no reason* to think that our moral beliefs are *true*.⁵⁴ The central thought behind evolutionary debunking

54 Vavova, "Debunking Evolutionary Debunking." In Vavova's terminology, the former argument is based on the principle *GOOD*, and the latter argument is based on the principle *NO GOOD*.

arguments is that evolution is an “off-track” influence, which would make it a *coincidence* if our beliefs were true. The Moral Closure Argument, on the other hand, is concerned not with *coincidental truth* or *off-track influences*, but with *evidential underdetermination*.⁵⁵

Furthermore, we can see that the Moral Closure Argument avoids the flaws in existing debunking arguments. According to the debunking argument advanced by Street, our moral beliefs are the product of influences that do not track the moral facts. Vavova argues in response that we have no good reason to think that these influences are “off-track” unless we have moral knowledge.⁵⁶ Without moral knowledge, says Vavova, we have no way of knowing what the moral facts are, and thus no way of knowing whether evolutionary influences are off-track or on-track. Since the Moral Closure Argument does not concern off-track influences, Vavova’s criticisms do not affect the Moral Closure Argument. Enoch criticizes Street’s debunking argument by arguing that it is no coincidence that our moral beliefs are true, because we can provide satisfying “third-factor” explanations of the reliability of our moral beliefs.⁵⁷ But again, because the Moral Closure Argument does not concern notions of coincidence or unexplainable reliability, Enoch’s response to Street’s debunking argument is a *non sequitur* from the perspective of the Moral Closure Argument.

Enoch and Vavova realize this; they both distinguish between evolutionary debunking arguments and closure arguments based on evidential underdetermination. Against evolutionary debunking arguments, they discuss knowledge of off-track influences and third-factor explanations. But against the Moral Closure Argument, they offer only the Special Case Objection. We have seen why the Special Case Objection fails.

8.2. *But in the Vicinity of Debunking*

One might worry that this defense against May’s objection is superficial. Not all debunking arguments appeal to the idea of coincidence, and not all debunking arguments are equally vulnerable to Vavova’s or Enoch’s objections. Thus, while debunking arguments typically appeal to a different array of epistemic concepts and principles than the Moral Closure Argument, one might suspect that the Moral Closure Argument is, on some level, an attempt to articulate *the same thought* that skeptics have been attempting to articulate by advancing evolutionary debunking arguments: that the availability of naturalistic explanations of our experiences supports moral skepticism.

55 Street, “A Darwinian Dilemma for Realist Theories of Value.”

56 Vavova, “Debunking Evolutionary Debunking.”

57 Enoch, *Taking Morality Seriously*.

If this is how we are to understand May's objection, then the charge has much more merit. The Moral Closure Argument is within a cluster of skeptical arguments influenced by Harman's famous discussion of moral observation and explanation.⁵⁸ Many different skeptical arguments are a part of this cluster. Arguments from disagreement argue that the best explanation of disagreement is that our moral beliefs are the product of cultural conditioning.⁵⁹ Evolutionary debunking arguments attempt to show that off-track evolutionary influences are the best explanation of our moral attitudes.⁶⁰ Parsimony arguments attempt to show that explanations that do not include moral facts are best in virtue of being simplest.⁶¹ There is a strong family resemblance between all of these arguments.⁶² All have something to do with the fact that we do not seem to need to include moral facts in the best explanations of our experiences. The fact that Error-ECP gives such a good explanation of our experiences is an essential aspect of the Moral Closure Argument. That puts the Moral Closure Argument squarely within this family of skeptical arguments.

But this is no objection to the Moral Closure Argument. While this family of skeptical concerns is familiar, every way of articulating the challenge thus far has been met with responses that anti-skeptics have, for the most part, been satisfied with.⁶³ Yet as I have been arguing here, the standard responses to the Moral Closure Argument are very weak. This makes the Moral Closure Argument a particularly potent member of this family of skeptical arguments, as we have yet to find a compelling response to it.

9. CONCLUSION

While the Moral Closure Argument could be viewed as a version of a number of familiar challenges, it constitutes a fruitful perspective on these challenges because it avoids the most substantial objections to other closure arguments (sections 3–6), and to evolutionary debunking arguments (section 8), without overgenerating skeptical results (section 7). There may still be effective ways to respond to the Moral Closure Argument. But if there are, they have not yet been

58 Harman, *The Nature of Morality*.

59 Mackie, *Ethics*, ch. 1; Wedgwood, "The Moral Evil Demons."

60 Street, "A Darwinian Dilemma for Realist Theories of Value."

61 Harman, *The Nature of Morality*.

62 Machuca, "Moral Skepticism."

63 Joyce, "Arguments from Moral Disagreement to Moral Skepticism."

identified. That is why the Moral Closure Argument deserves to be taken seriously by anti-skeptics.

Wuhan University
matlutz326@gmail.com

REFERENCES

- Alston, William P. "Epistemic Circularity." *Philosophy and Phenomenological Research* 47, no. 1 (September 1986): 1–30.
- Bedke, Matthew. "Against Normative Naturalism." *Australasian Journal of Philosophy* 90, no. 1 (2012): 111–29.
- BonJour, Laurence, and Ernest Sosa. *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Maldwell, MA: Blackwell Publishing, 2003.
- Button, Tim. "Deflationary Metaphysics and Ordinary Language." *Synthese* 197, no. 1 (January 2020): 33–57.
- Cohen, Stewart. "Justification and Truth." *Philosophical Studies* 46, no. 3 (November 1984): 279–95.
- Conee, Earl, and Richard Feldman. "Replies." In *Evidentialism and Its Discontents*, edited by Trent Dougherty, 283–323. Oxford: Oxford University Press, 2011.
- Dougherty, Trent, and Patrick Rysiew. "Experience First." In Steup, Turri, and Sosa, *Contemporary Debates in Epistemology*, 17–21.
- Dretske, Fred. "Epistemic Operators." *Journal of Philosophy* 67, no. 24 (December 1970): 1007–23.
- Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2011.
- Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press, 1977.
- Hawthorne, John. "The Case for Closure." In Steup, Turri, and Sosa, *Contemporary Debates in Epistemology*, 40–55.
- Horgan, Terry. "Troubles for Bayesian Formal Epistemology." *Res Philosophica* 94, no. 2 (April 2017): 233–55.
- Huemer, Michael. *Ethical Intuitionism*. Houndmills, UK: Palgrave Macmillan, 2005.
- . *Skepticism and the Veil of Perception*. London: Rowman and Littlefield, 2001.
- Joyce, Richard. "Arguments from Moral Disagreement to Moral Skepticism."

- In *Moral Skepticism: New Essays*, edited by Diego E. Machuca, 141–62. New York: Routledge, 2018.
- . *The Evolution of Morality*. Cambridge, MA: MIT Press, 2007.
- . “The Skeptick’s Tale.” *Philosophy and Phenomenological Research* 78, no. 1 (January 2009): 213–21.
- King, Jeffrey C., Scott Soames, and Jeff Speaks. *New Thinking about Propositions*. Oxford: Oxford University Press, 2014.
- Kulp, Christopher. “Moral Facts and the Centrality of Intuitions.” In *The New Intuitionism*, edited by Jill Graper Hernandez, 48–66. London: Continuum International, 2011.
- Lutz, Matt. “Defusing the Counterinduction Parody.” *Philosophia* 46, no. 2 (June 2018): 379–85.
- . “What Makes Evolution a Defeater?” *Erkenntnis* 83, no. 6 (December 2018): 1105–26.
- Machuca, Diego E. “Moral Skepticism: An Introduction and Overview.” In *Moral Skepticism: New Essays*, edited by Diego E. Machuca, 1–2. New York: Routledge, 2018.
- Mackie, John. *Ethics: Inventing Right and Wrong*. London: Penguin UK, 1977.
- Majors, Brad. “Moral Explanation.” *Philosophy Compass* 2, no. 1 (January 2007): 1–15.
- May, Jonathan. “Skeptical Hypotheses and Moral Skepticism.” *Canadian Journal of Philosophy* 43, no. 3 (2013): 341–59.
- McCain, Kevin. *Evidentialism and Epistemic Justification*. Abingdon, UK: Routledge, 2014.
- McCann, Hugh. “Conative Intuitionism.” In *The New Intuitionism*, edited by Jill Graper Hernandez, 29–47. London: Continuum International, 2011.
- McPherson, Tristram. “Moorean Arguments and Moral Revisionism.” *Journal of Ethics and Social Philosophy* 3, no. 2 (June 2009): 1–25.
- Moore, G. E. *Principia Ethica*. New York: Dover Publications, 1903.
- Oddie, Graham. *Value, Reality, and Desire*. Oxford: Oxford University Press, 2005.
- Poston, Ted. *Reason and Explanation*. New York: Palgrave Macmillan, 2014.
- Pryor, James. “The Skeptic and the Dogmatist.” *Noûs* 34, no. 4 (December 2000): 517–49.
- . “What’s Wrong with Moore’s Argument?” *Philosophical Issues* 14, no. 1 (October 2004): 349–78.
- Railton, Peter. “Moral Realism.” *Philosophical Review* 95, no. 2 (April 1986): 163–207.
- Rosen, Gideon. “What Is Normative Necessity?” In *Metaphysics, Meaning, and*

- Modality: Themes from Kit Fine*, edited by Mircea Dumitru, 205–33. Oxford: Oxford University Press, 2020.
- Russell, Bertrand. *The Problems of Philosophy*. Oxford: Oxford University Press, 2001.
- Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford University Press, 2003.
- Sinnott-Armstrong, Walter. *Moral Scepticisms*. Oxford: Oxford University Press, 2006.
- Star, Daniel. “Moral Knowledge, Epistemic Externalism, and Intuitionism.” *Ratio* 21, no. 3 (September 2008): 329–43.
- Street, Sharon. “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127, no. 1 (January 2006): 109–66.
- Steup, Matthias, John Turri, and Ernest Sosa, eds. *Contemporary Debates in Epistemology*, 2nd ed. Malden, MA: Wiley Blackwell, 2014.
- Stroud, Barry. *The Significance of Philosophical Scepticism*. Oxford: Oxford University Press, 1984.
- Sturgeon, Nicholas. “Moral Explanations.” In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 229–55. Ithaca, NY: Cornell University Press, 1988.
- Thomasson, Amie. *Ontology Made Easy*. Oxford: Oxford University Press, 2014.
- . *Ordinary Objects*. Oxford: Oxford University Press, 2007.
- Van Cleve, James. “Reliability, Justification, and the Problem of Induction.” *Midwest Studies in Philosophy* 9, no. 1 (September 1984): 555–67.
- Vavova, Katia. “Debunking Evolutionary Debunking.” In *Oxford Studies in Metaethics*, vol. 6, edited by Russ Shafer-Landau, 76–101. Oxford: Oxford University Press, 2016.
- Vogel, Jonathan. “Cartesian Skepticism and Inference to the Best Explanation.” *Journal of Philosophy* 87, no. 11 (November 1990): 658–66.
- Wedgwood, Ralph. “Internalism Explained.” *Philosophy and Phenomenological Research* 65, no. 2 (September 2002): 349–69.
- . “The Moral Evil Demons.” In *Disagreement*, edited by Richard Feldman and Ted A. Warfield, 216–46. Oxford: Oxford University Press, 2010.
- Williamson, Timothy. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.
- . “Knowledge First.” In Steup, Turri, and Sosa, *Contemporary Debates in Epistemology*, 1–9.

JOURNAL of ETHICS & SOCIAL PHILOSOPHY
<http://www.jesp.org>
ISSN 1559-3061

The *Journal of Ethics and Social Philosophy* (JESP) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes.



Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.