JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXXI · NUMBER 1
February 2026

ARTICLES

DISCUSSION

OPEN ACCESS FREE FOR AUTHORS AND READERS ISSN 1559-3061

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XXXI · NUMBER 1
February 2026

ARTICLES

DISCUSSION

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 Division of Arts and Humanities at New York University Abu Dhabi.

JESP 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 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. *JESP* also endorses and abides by the Barcelona Principles for a Globally Inclusive Philosophy, which seek to address the structural inequality between native and nonnative English speakers in academic philosophy.

JESP publishes articles, discussion notes, and occasional symposia. Articles normally do not exceed 12,000 words (including notes and references). *JESP* sometimes publishes longer articles, but submissions over 12,000 words are evaluated according to a proportionally higher standard. Discussion notes, which need not engage with work that was published in *JESP*, should not exceed 3,000 words (including notes and references). *JESP* does not publish book reviews.

Papers are published in PDF format at https://www.jesp.org. All published papers receive a permanent DOI and are archived both internally and externally.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.4432

WHAT FICTIONALISTS GET WRONG ABOUT THE VALUE OF WINNING

William J. Morgan

THE SHEER NUMBER of fans who watch sports and cheer on their favorite individual athletes and teams is astonishing. For instance, global television broadcasts for the summer Olympic Games typically attract over three billion people, while the television audience of the World Cup in football typically ranges from three to five billion people. The national audience for major sporting events is no less astonishing: the US television audience for the 2024 summer Olympic Games was over thirty million, and that for the 2024 Super Bowl was close to a 124 million.[1] But it is not just the sheer number of fans who watch and follow these sports that is eye opening but rather the typical passionate fervor in which they watch and follow them as well. Indeed, they often act as if their very lives depend upon whether the players or teams they root for come out on top.

But the wide following that such games attract and the passionate responses of their fans they induce pose a puzzle. For it is puzzling to many why enthusiasts of sports get so emotionally caught up and absorbed in the outcome of games in spite of the fact that those outcomes rarely if ever have any appreciable bearing on their ordinary lives. In short, there seems to be a radical disconnect between fans' over-the-top emotional engagement in whether or not their favorite players or teams succeed and the scant value of winning itself. It might be tempting to write off the masses who flock to such games as simply irrational, as going all in for game outcomes that have little if any value to speak of—outcomes that in the larger scheme of things, simply do not matter. Surely, however, simply dismissing what such an impressively large number of people take with utmost seriousness and passionately follow is the height of intellectual arrogance.

Fortunately, some philosophers at least have taken the puzzle of sports seriously. The philosophers I have particularly in mind follow Kendall Walton's lead in his article "It's Only a Game!"[2] In it, Walton suggests that make-believe is

1 Figures are from the Roadtrips webpage "Sports Events," https://www.roadtrips.com (accessed November 6, 2025).

2 Walton, "It's Only a Game!" Hereafter, this article is cited parenthetically. See also Baron-Schmitt, "Who Cares About Winning?"; Borge, *The Philosophy of Football*; and Moore,

what accounts for fans' preoccupation with winning. I call such philosophers *fictionalists* because the case they make for the make-believe character of sports fandom rests on an analogy between the puzzle of fiction and the aforementioned puzzle of sports.

The puzzle of fiction concerns how it is that enthusiasts of fictional films, stories, and plays manage to get emotionally caught up in characters who do not exist and made-up stories about them. The explanation, fictionalists suggest, is, roughly, that enthusiasts pretend that such characters exist and that the stories about them are real (76). By make-believing that these fictional characters actually exist—that, for example, Willy Loman is a real person or that the menacing green slime on the screen is a real monster—audience members are induced to emotionally respond to them in ways that are phenomenologically similar to how they emotionally respond to events in real life. They are able, as it were, to "feel" for Willy Loman's unenviable plight, to pity him, or to "feel" that the slime poses a danger to them and thus to fear it. However, they know full well that Willy Loman does not exist and that the slime is merely an image projected on the screen. So, Walton claims, they do not "literally" pity Willy Loman but only "quasi-pity" him, and they do not "literally" fear the slime but only "quasi-fear" it. It is only fictionally true then that audiences pity Loman and fear the slime. Nevertheless, Walton insists, their pretense-generated experiences touch off genuine feelings, for example, of distress or of being emotionally unsettled by what they observe, even if they do not literally pity or literally fear these fictional entities.[3]

Fictionalists claim that a similar flight of imagination explains why sports fans care so much about game outcomes. In this case, however, what sports fans make believe is not, of course, that the players or teams they follow exist or that what happens to them is real but rather that winning by the players or teams they follow is much more important and valuable than they know it really is. If the explanations that fictionalists give to both puzzles is persuasive, then enthusiasts of the fictional arts and sports should need no reminding or reassuring that, as Walton writes, "It's just a story; it's just a game" (77). Nor

"Do You Really Hate Tom Brady?"

3 The latter, what Walton calls the "positive" thesis of his make-believe theory—that audience members are genuinely emotionally moved, distressed, unsettled, or uplifted by what they may observe in a fictional play or watch in a film—is meant to complement his "negative" thesis—that they do not literally pity, fear, grieve, or admire fictional characters. Too many critics, he laments, have glommed on to his negative thesis, to the detriment of his positive thesis. As a result, he argues, they fail to see that his "make-believe theory was designed to help explain our emotional responses to fiction, not to call their very existence into question" ("Spelunking, Simulation, and Slime," 275).

should passionate followers of sports worry, Walton further insists, that being a sports fan is irrational. The idea that fans temporarily lose their minds while watching sports, he insists, "is no more attractive than the idea readers of a story lose their senses, temporarily, and believe in goblins or magic rings" (77).

What is notable about the fictionalist solution to the puzzle of sports is that it defends the rationality of being a fan while conceding that the object of a fan's great affection and care—winning—has no real value to speak of. Indeed, it is only because fictionalists reject that winning has any independent importance or value as an end in itself that they argue it can acquire such (fictional) importance and value only if observers pretend that it does. What is notable if not ingenious about the fictionalist solution, however, is what proves to be its undoing. For their main claim that being a fan is rational so long as we imaginatively inflate the value of winning founders on the very largely unargued supposition on which it is based: that winning is a valueless end. I argue, contrarily, that there is a very strong case to be made that winning is a valuable end, and that value has to do with the significant achievement that winning, rightly understood, represents. In my view, what is rational about being a fan is inextricably connected to the value that winning possesses as a form of human achievement, and therefore, it is that achievement value that explains and justifies why fans care so much about game outcomes. If I am right about this, then there is nothing irrational or odd about why those who watch sports fixate on game outcomes, nor *a fortiori* why those who play them try mightily to win—and so there is no puzzle that needs solving in fictionalist or other fanciful theoretical terms.

I

It is important from the outset to get clear about both what sorts of competitive games and sports and what kind of spectators and players are captured by the so-called puzzle of sports. The examples that fictionalists provide offer a helpful start in this direction, but as we see below, they prove highly misleading in one important respect.

The main examples that fictionalists entertain of spectators' emotional engagement in games all involve top-level sports in which the players featured are highly competent, expert performers.[4] In this vein, Nathaniel Baron-Schmitt asks us to consider a hypothetical fan named Dan who is "on the edge of his seat" as he intently watches his favorite professional American football team, the

4 As we see below, that spectators mostly root for expert, highly competent players and teams is crucial to my ensuing argument.

Dallas Cowboys, on television.[5] Similarly, Walton poses the example of real-life Boston Red Sox baseball fans "scream[ing] their hearts out during the game" as they urge their team to victory (76). Finally, Joseph Moore calls attention to his own exuberant behavior as a fan watching Mikaela Schriffen's stunning performance in a 2016 world alpine skiing competition as he and those around him ecstatically rejoice in her impressive victory.[6]

What we can glean from these examples of sports fans that in my estimation point us in the right direction in getting a grip on the puzzle of sports is, first of all, that the objects of fans' highly charged emotional engagement are, one and all, complex, sophisticated competitive games and sports such as football and baseball. The second thing we can glean from them is that in virtue of the in-built formidable structural challenges that such sports pose, the focus of fans' intense attention is largely on highly competent, expert players who have what it takes in the way of skill, ingenuity, and finesse, not to mention dedication to their crafts, to succeed at them.

Fictionalists thus hold that it is mainly sophisticated competitive games and sports and the corresponding highly accomplished and committed players that play them so well that draw spectators in the exceedingly large numbers they do and generate the emotional fervor characteristic of fans who take them in. So when fictionalists talk about the over-the-top emotional way in which fans root for the success of their favorite players or teams, they are not talking about simple, rudimentary children's games such as tic-tac-toe or rock-paper-scissors—which provide players only a limited number of choices and possible strategies to pursue, require little in the way of skill or ingenuity, and are practically effortless to play—nor games of chance—which require no skill or ingenuity to play at all. That games of the latter two kinds do not figure in what fictionalists call the puzzle of sports is further borne out by empirical evidence. After all, children's games and games of chance generate little to no fan interest to speak of, let alone the kind of fevered emotional behavior typical of fans of sophisticated games like baseball.

However, the kind of fans that fictionalists showcase in their examples to shed light on this facet of the puzzle of sports is itself, alas, puzzling. That is because they allege that these fans hold oddly schizophrenic attitudes about game outcomes. On the one hand, fictionalists maintain that such fans care intensely while a game is in progress about how their favorite players or teams fare against their opponents and whether they will ultimately prevail over them. On the other hand, they maintain that as soon as the game is over, these same

5 Baron-Schmitt, "Who Cares About Winning?" 250.

6 Moore, "Do You Really Hate Tom Brady?" 245.

fans do not care one whit whether their favorite players or teams won or lost, if, that is, they even remember any longer whether they succeeded or failed.[7] For example, Walton avers that "many" fans typically get emotionally carried away during the game but "forget the game quickly after it is over, much too quickly for people who care as much as they seem to during the game" (77). That is why he claims that the moment the game ends, these "avid" sports fans go back to living their lives "as though nothing much had happened—even if the home team suffered a devastating and humiliating defeat" (77). Similarly, Baron-Schmitt claims that Daniel "is on the edge of his seat" while the game is in progress but "will not spend any energy tomorrow lamenting a Cowboys defeat."[8] And Moore describes his own experience of being "overwhelmed with what can only be described as . . . euphoria" over Schriffin's remarkable victory but confesses that his euphoria was short-lived, lasting, in his own words, "all of fifteen seconds, as my attention quickly shifts to beating the freezing crowd to the warm shuttle-bus that will take me back to my car."[9]

What seems amiss about the odd attitudinal profile of the sports fans that fictionalists target in this regard, about what Baron-Schmitt aptly describes as their "casual but enthusiastic" outlook, is that most casual fans are anything but enthusiastic about how the players or teams they follow are faring during the game and about whether they succeeded or failed after the game; conversely, most enthusiastic fans are anything but casual about how the players or teams they follow are faring during the game and whether they succeeded or failed after the game. In fact, the anecdotal, observational, and statistical evidence shows convincingly that most fans who follow top-level sports fall squarely into the enthusiastic category, and the largest bloc of these enthusiastic fans is made up of partisan fans.[10] It is these partisan enthusiastic types, I suggest, that are the kind of fans most disposed to exhibit the emotionally high-strung and overwrought behavior that fictionalists like Walton write about while a game is ongoing and the most disposed, *contra* fictionalist claims to the contrary, to celebrate and

7 Baron-Schmitt, "Who Cares About Winning?" 250.

8 Baron-Schmitt, "Who Cares About Winning?" 250.

9 Moore, "Do You Really Hate Tom Brady?" 245. In this same passage, Moore notes, "Not only is my euphoria ephemeral, but I have no inclination to act on it by, say, congratulating Mikaela."

10 According to a 2023 American Sports Fanship Survey, 47 percent of Americans are "avid" or "involved" sports fans whose attitudes about winning nicely dovetail with those I call partisan enthusiastic fans. Twenty-six percent of those surveyed are described as "casual" fans, and the remaining 27 percent, who watch and follow sports only sporadically, are considered "nonfans." The survey was conducted by the Siena College Research Institute and the Jandoli School of Communication at St. Bonaventure University. See https://sri.siena.edu/american-sports-fanship-survey-2023/.

revel in the successes of their favorite players and team or, as the case may be, to bemoan and despair their losses long after a game has concluded.[11]

Where the fictionalist account of sports fans goes wrong, I suggest, is the apparent melding of what are ostensibly two distinct categories of casual and (partisan) enthusiastic fans with hard-to-miss clashing attitudes about game outcomes into one phenomenologically peculiar and discordant category of the casual but enthusiastic fan. Why I think this ranks as a misstep is because there are precious few fans for whom this odd bivalent attitudinal profile is apposite—who at one moment are enthralled by the prospect that the players or teams they follow are on the cusp of winning and who mere moments later are utterly indifferent as to whether those same players or teams won or not. In short, most sports fans are just not this fickle when it comes to their attitudes about winning. Why I think this is a significant misstep on the part of fictionalists is because if their make-believe theory of sports fandom is to have any explanatory force or critical bite, it must account for a sufficiently large and representative sample of the emotional behavior of actual sports fans. Fickle sports fans are anything but such an inclusive, representative sample. That is not to say that there are no fans that have such fleeting and radically shifting attitudes about game outcomes.[12] Indeed, given the exceedingly large number of fans who go in for competitive games and sports, there are bound to be some fans who fit this curious attitudinal profile. At most, however, if the empirical evidence is to be believed, they comprise a tiny group of fringe outliers.[13]

11 The markedly different attitudinal profile of partisan enthusiastic fans is nicely illustrated by Kadlac's description of his own partisan allegiance to the football team Tottenham. As he writes, "when I watch Tottenham play, I care about the outcome of the game. This concern is not a mild preference in the way I might prefer … an apple fritter rather than a glazed donut." Rather, whether Tottenham wins or loses, he continues, "will affect my emotions in the future," quite unlike whether his preference for an apple fritter goes unsatisfied because none are available. Such unfulfilled mild preferences, he notes, are quickly forgotten. "By contrast, if Tottenham wins, my mood will be lifted, not only in the immediate aftermath of the victory, but every time I think about that occasion. And should they lose, I will be disappointed—rolling over the various missed opportunities and mistakes in my head" (*The Ethics of Sport Fandom*, 57). In fact, many such fans remember the high moments of victories over heated rival players and teams, not to mention the bitter defeats, years after their occurrence. Walton's claim in that "many" fans quickly forget even "devastating and humiliating defeat[s]" of the home team does not stack up against the available empirical evidence ("It's Only a Game!" 77).

12 That is why I see no reason to doubt Moore's fictionalist description of himself as just such a fickle fan.

13 This is why I pay no mind to the other main argument that fictionalists make to show that fans only quasi-care about winning that zeroes in on the so-called fickle fan's on-again, off-again outlook on game outcomes.

To be clear, however, in calling into question the bona fides of a fickle fan as a reputable/representative sports fan type, I do not call into question the bona fides of a casual fan as a reputable/representative sports fan type. As already indicated, there is nothing out of the ordinary about the attitudinal profile of casual fans who, unlike fickle fans, are consistently blasé about who is winning both during and after the game, save occasional and brief frissons of emotion, and who, as noted, make up around 25 percent of the sports fan base. More importantly for my purposes, however, is that because casual fans are typically not greatly invested, emotionally or otherwise, in games or their outcomes, there is no reason to suppose that they need to rely on some sort of imaginative assist to engage with games. For casual fans not only do not *believe* winning matters but also do not generally *behave* like it matters.

However, if I am right that it is (partisan) enthusiastic fans' emotionally exaggerated obsession with winning that in part drives the notion that there is such a thing as the puzzle of sports, then there is at least a *prima facie* case to be made by fictionalists that such fans are incapable of such emotionally charged behavior unless they imaginatively inflate the value and importance of winning. As we have seen, this is precisely the claim that fictionalists advance by insisting that winning has little to no value or importance in itself and by insisting that for the great majority of fans, whether their favorite player or team wins has no appreciable bearing on the rest of their lives. It is in this respect that fictionalists suggest that sports fans' pretense-induced quasi-caring about game outcomes is "tantalizingly" similar to a playgoer's pretense-induced quasi-pity of Willy Loman and a filmgoer's quasi-fear of a slime monster. In the former case, of course, as previously remarked, the object of fans' make-believe is not the all too real players nor the all too real actions they take to win. Furthermore, in competitive games and sports, there is no script that has already been plotted out by some controlling author in the background, which the players faithfully follow and the fans are prescribed to respond to accordingly. Notwithstanding these obvious differences, however, fictionalists argue that it is the make-believe that winning matters that is the key to understanding spectators' remarkable emotional investment in game outcomes. As Walton writes, "superimposed on a modest genuine interest in the outcome, there is, frequently, a pretense of much greater concern" that does not come close to the actual value of that outcome (78). Walton further argues this modest interest in winning and the corresponding imaginative magnification of it do not simply coexist but rather reinforce one another. In his view, the feelings of excitement and pleasure that accompany this modest interest in winning serve as what he calls "props" in the make-believe amplification of its importance and value (78). It is therefore by make-believing that winning matters that fictionalists like Walton maintain spectators come to believe that it matters.

However, fictionalists qualify their make-believe account in one narrow respect. They concede that for *some* fans, winning really does matter if it instrumentally benefits them in some tangible way (78). For instance, winning very much matters to spectators who bet on sports and/or whose personal well-being is otherwise directly impacted. Such spectators, Walton and company acknowledge, need no imaginary assist to take winning seriously. Nonetheless, they hasten to point out, only a tiny fraction of sports fans stands to benefit from game outcomes materially or otherwise. So the exception they carve out for such fans does not, in their view, undermine or weaken their main thesis that for *most* or *many* fans, pretense is what lies behind the great importance and value they ascribe to winning.

The upshot of the fictionalist argument is thus that sports fans' intense emotional investment in game outcomes is a figment of their imagination. What is really going on, as fictionalists see it, is that many fans merely pick a player or team to root for arbitrarily or on a whim, perhaps because they take a fancy to their uniforms, perhaps because their friends are supporters of a particular player or team, or for countless other capricious factors (76). This inconsequential, modest, acquired interest in their favorite players' or teams' wins or losses is then bootstrapped by their make-believing that they have good reasons for their obsession over their successes and failures. Add to this, the pageantry that often accompany such games—the colorful banners, team logos, large numbers of people cheering in unison, which all serve as props to fuel the fans' imagination and fictional belief in the importance of these events—and we have all we need, according to fictionalists, to account for the striking mismatch between the remarkable enthusiasm that fans display in anticipation of its outcome and the utter triviality of that outcome. After all, to quote Walton once again, "it's only a game."[14]

14 Generally speaking, what fictionalists have to say about spectators' pretense that winning matters goes for players as well. However, they maintain that unlike actors on a stage, those who play games are mostly unaware of the pretense that lies behind their imaginary inflation of winning. See Borge, *The Philosophy of Football*, 207. That is because, as Walton writes, "it will not always be obvious whether and to what extent a competitor or spectator engages in make-believe. . . . We must [therefore] . . . leave room for tacit, implicit pretense, pretense that may not be evident even to the pretender" ("It's Only a Game!" 82–83). Fictionalists also concede, just as they do for spectators (e.g., who bet on games), that winning really does matter to players who financially profit from their gameplay or derive prestige and fame from it. Since my focus is on spectators in this article, I confine whatever I have to say about players in this regard to how their gameplay importantly figures in the achievement value that, I argue, lies behind spectators' fixation on winning.

II

My critique of the fictionalist make-believe theory of sports fans' extraordinary emotional investment in game outcomes zeroes in on their failure to see that there are compelling reasons to think that it is the intrinsic achievement value of winning, properly understood, that lies behind and justifies fans' fixation on winning. But before I try to make that case, I must first clarify the conception of achievement that I have in mind here.

What I mean by *achievement* is mostly what its main proponents mean by it.[15] First and foremost, as they see it, what distinguishes achievements from nonachievements is the "difficulty" involved in accomplishing the various goals and ends that govern them. Achievements are thus not a simple matter of accomplishing one's ends whatever they might be. For example, getting dressed in the morning is for most adult, able-bodied people in otherwise ordinary situations not an achievement in the relevant sense. Rather, it is only activities in which accomplishing one's ends is difficult to do that one can rightly be said to have accomplished something. By *difficulty*, Thomas Hurka primarily means how complex and challenging those activities prove to be.[16] To rank as a genuine achievement, therefore, the activity must involve the exercise of sophisticated skills, ingenuity, and practical reasoning. The higher degree of those skills, the more varied the skills and the more complex steps involved in their execution, the greater the achievement. It also follows that accomplishing difficult (complex) goals requires intense effort, which is not simply a matter of how much total effort is expended but how much "*effort over a certain level of intensity*" is expended.[17] Finally, if difficulty is the signature feature of achievement, it follows that a low probability of success in achieving an end is a further characteristic feature of achievement.[18]

The above three features spell out what counts as an achievement. What, though, is to be said about the value of achievement? My answer is again the same perfectionist answer that Hurka and company give to the value of

15 See Hurka, "Games and the Good"; Bradford, *Achievement*; and von Kriegstein, "On Being Difficult." I qualify my claim that I *mostly* follow what theorists of achievement have to say about it for two reasons. First, I do tweak some of the features they designate as necessary conditions for an activity to be an achievement to accommodate my focus on sports. Second and more importantly, I argue that one further necessary feature of what counts as an achievement that they do not mention is that the accomplishment of the goal of the activity in question must be *socially recognized* as an important and valuable goal in its own right if it is to be regarded as a genuine achievement.

16 Hurka, "Games and the Good," 221.

17 Bradford, *Achievement*, 49.

18 Von Kreigstein, "On Being Difficult," 55.

achievement rightly understood: the value of different kinds of achievement has to do with the exercise of those basic capacities that make us the distinctive human agents that we are. On this perfectionist view, the excellent exercise of those basic human capacities is intrinsically valuable.[19] Bradford specifies two epistemic criteria for identifying what counts as basic human capacities. According to the first, basic human capacities are those whose exercise is "near universal" or "near inevitable."[20] As George Sher puts it, basic human capacities are ones that almost "all human beings possess, and … must be one[s] whose exercise [their] possessors either can't avoid at all, or else can avoid intermittently."[21] According to the second, a basic human capacity is one that is "intuitively worth developing," and this idea derives from the fact that it is the exercise of those basic capacities that "make[s] us good as human beings."[22] It thus follows that we should intuitively regard the development of those capacities as worthwhile and valuable for their own sake.[23]

The kinds of achievement that satisfy the three criteria specified above make up a special class of achievement—the sort that Bradford aptly calls "capital A achievements."[24] They are special because of their exceptional character, because of the human excellences that are required to accomplish them. These are the kinds of achievement that evoke the awe and admiration of those who bear witness to them enriching their own lives in the process. What Bradford and company have in mind when they write about such especially significant achievements are, to name just a few, writing a scholarly paper that gets published in a well-respected peer-reviewed journal, having one's paintings shown in a major gallery, writing a novel that is awarded a reputable literary prize, and so on.

Sports and the goal of winning in sports, I argue, belong in this special class as paradigmatic examples of stellar human achievement. But as noted, fictionalists claim otherwise because they regard winning to be a trivial goal that lacks the value—not to mention the complexity—that achievement theorists maintain is a signature feature of genuine achievements. As fictionalists like Baron-Schmitt see it, winning a game is an "odd, unimportant" one in which players try to get things like "an orange sphere … through a hoop" that

19 Bradford, *Achievement*, 115.

20 Bradford, *Achievement*, 116.

21 Sher, *Beyond Neutrality*, 199 (quoted in Bradford, *Achievement*, 116).

22 Bradford, *Achievement*, 117.

23 Bradford, *Achievement*, 116–17.

24 Bradford, *Achievement*, 4.

is weirdly suspended ten feet from the ground.[25] If one team manages to get an orange sphere through a hoop more times than the other, they are crowned the winner. Winning in other sports presumably can be similarly described in these same flat, austere physical terms. For example, winning a footrace can be described as simply crossing the finish line first, and winning a golf game can be described as getting a small ball into several small holes in the ground. What value can accomplishing such states of affairs possibly have, fictionalists incredulously ask, and what is especially difficult about getting a small ball into a small hole, they further incredulously ask. For if winning amounts to no more than getting orange spheres through hoops and the like, then as a standalone end, fictionalists find it inexplicable how humans could find value in such outcomes, much less care so greatly about them. Unless, of course, as fictionalists stoutly maintain, players and fans alike simply make believe that getting orange spheres through hoops really matters when they know it obviously does not.

The problem, however, is that fictionalists have misdescribed winning by confusing what Bernard Suits's in his classic account of sports calls the *prelusory* goal of sports with what he calls the *lusory* goal of sports.[26] By prelusory goal, Suits means the "specific achievable states of affairs" that players strive to achieve. The prelusory goal of basketball is of course precisely what Baron-Schmitt wrongly describes as winning a basketball game: one team gets an orange sphere more times through the hoop than the opposition. Or, as one would similarly and wrongly describe winning a footrace: crossing a finish line first. Suits calls these goals prelusory because they can be "described before, or independently of, any game in which [they] may be, or come to be a part."[27] It is because prelusory goals specify only the states of affairs players aim to accomplish that they can be detached from the means and manner by which they are realized, which is why it is possible to attain them aside from the games in which they figure—this is of course what cheaters do when they violate the rules in order to best their opponents.[28] What is more, as standalone ends, they are not by any stretch of the imagination hard to accomplish, since one can get an orange sphere through a hoop more times than someone else simply by using a ladder to stuff the ball through the hoop or can get a small ball into a hole in the ground by simply using one's hand to drop it in. By contrast, winning demands that players do more than simply achieve a specific state of

25 Baron-Schmitt, "Who Cares About Winning?" 248.

26 Suits, "The Elements of Sport." 40.

27 Suits, "The Elements of Sport," 40.

28 As Suits writes, when players break the central rules of a game, they "fail … to play the game at all" ("The Elements of Sport," 41).

affairs, which is why, as noted, Suits calls these the lusory goals of sports. For to realize the lusory goal of a sport, one must not only get an orange sphere through the hoop more times than the opposition or get a ball into holes in the ground but do so using only the means permitted by the rules. What Suits calls the *constitutive rules* of sports make the attainment of these otherwise simple and uncomplicated prelusory goals much more difficult and complex by prohibiting players from using more efficient means in favor of less efficient means. For example, in basketball, it is useful but prohibited to use a ladder to get the orange sphere through the hoop; similarly, it is useful but prohibited to use laser-guided balls in golf. So to win a basketball game, one must, among other things, get the orange sphere (the ball) through the hoop by launching it with one's hands without taking too many steps and shoving one's opponent aside; and to win a golf game, one must, among other things, stand at some considerable distance from the intended target and use a golf club to propel the ball into the hole in the ground. What for Suits so crucially distinguishes lusory goals from prelusory, standalone ones, then, is that the specification of the rules, means, and manners by which these states of affairs must be accomplished are themselves *intrinsic features of lusory goals*.[29]

29 Baron-Schmitt's misdescription of winning is owed, I suggest, to the abstract manner in which he describes it, for he chooses to describe it by adopting the vantage point of an alien observer of Earth he calls Joe, who is trying both to understand what motivates earthlings to do the things they do and to make sense of what they do. Joe, he claims, is able to make sense of most of the things earthlings do but is flummoxed by why they care about "odd and unimportant things" like getting orange spheres through hoops. But I think Joe's bewilderment about such game outcomes has to do with the fact that from this maximally detached viewpoint, all he has to go on are the sheer physical actions that players execute, which seem to have something mystifyingly to do with the goal they are trying to realize. What such a detached observer has no way of accounting for are the social features that overlay these brute actions, which are conferred, as we have seen, by the rules that make the players' actions intelligible and give winning the normative force it has for them. These social features are the very ones Suits relies on to draw his distinction between the prelusory and lusory goals of sport. John Searle cautions us not to describe such social or what he calls institutional features of things like games in "purely brute terms" (*Speech Acts*). To illustrate his point, he imagines a group of trained observers who are tasked with describing American football using only brute facts. Unsurprisingly, all they are able to come up with, he writes, is that "at statistically regular intervals, organisms in like-colored shirts cluster together in roughly circular fashion (the huddle) . . . followed by linear clustering (the teams line up for a play), followed by linear interpenetration" (52). Searle rightly notes that what they thus describe is not American football because they leave out institutional features such as touchdowns, offside, points, first downs, etc., and all the true statements about football that one can make using these concepts. In Joe's case, he has no way of accessing such institutional features of sports like football because of his detached vantage point. And it is safe to say that from this same lofty standpoint, Joe would

It is thus because fictionalists like Baron-Schmitt misdescribe winning that they run together the prelusory goals of sports with their lusory goals and, as a consequence, find it unfathomable why players pursue victory so wholeheartedly and why fans care so greatly about the success of their favorite teams. Once we correct for this mistake and rightly understand winning as the accomplishment of a *lusory* goal, which is an estimable goal that demands the great skill, effort, and ingenuity that Hurka and company argue are the marks of "capital A," genuine achievements, any perplexity over players' ardent pursuit of victory and fans' ardent, passionate response to their winning ways should vanish.

I should make clear, however, that in identifying winning as the locus of the intrinsic value of achievement in these complex competitive games and sports, I do not claim that *only* those players or teams that prevail over their opponents in such competitions can be said to have genuinely achieved something. I thus acknowledge that those players and teams that play well against evenly matched opponents and yet lose have achieved something of intrinsic value as well. Being able to give one's opponents a good, well-played game regardless of the outcome ranks as a notable achievement on my account. What my focus on winning does suggest, however, is two things. First, while giving one's opponent a good game is an achievement, it is a lesser one than winning, which is, after all, the goal of athletic contests. Second, if players or teams are able to play just well enough to push their opponents to excel but are never able to actually defeat them, then there is good reason to regard their lesser achievements in this regard as appreciably diminished, if not, in fact, over time, as failed non-achievements. This once again suggests that winning is an ineradicable feature of what counts as an achievement in the realm of sports and games like them.

As a first step in better understanding how the achievement value of winning figures in fans' remarkable interest in and enthusiasm for such game outcomes, it is important to note an important difference in the relationship that obtains between producers (artists) of fictional works and their intended audiences and that between players' performances and the audiences they attract. In the case of fictional works, artists churn out theatrical plays, films, and literary works that not only *depict* some fictional series of events but crucially *prescribe* a point of view, a perspective, on how their intended audiences should respond to these imaginary scenarios.[30] We have already noted how this works in the case of horror films, in which those responsible for their production create fictional characters and fictional situations that directly aim to elicit responses

be at a loss, *contra* Baron-Schmitt, to understand what motivates and gives sense to almost everything earthlings do in their ordinary lives with only these brute facts at his disposal.

30 Gaut, *Art, Emotion, and Ethics*, 230.

of fear and terror from their target audiences. The success or failure of their efforts in this regard completely hangs on whether they in fact get the audience to respond accordingly. The relationship between athletes' performances and their audiences is not like this at all. On the contrary, the actions that players take in a game are entirely devoted to winning rather than to prompting their audiences to respond in some particular way. In fact, nothing the players do in the game needs to and rarely does consider how spectators might react to or be affected by their actions. This is why those who play top-level sports, unlike their fictional counterparts, can and do readily accomplish what they set out to do even if there is nobody watching—notwithstanding the fact that it is precisely these expert players who draw the attention and interest of the greatest number of enthusiastic fans. As Walton correctly notes, "no one arranges the events of the game to best advantage for appreciation [by spectators]—at least no one is supposed to" (81).

What Walton is getting at is that while sports are staged affairs, unlike fictional works, they are not scripted ones. That means that although sports fans bear witness to what excellences are achieved in the game, they are not the intended targets of those excellences. Sports are of course well suited to such bearing witness on the part of spectators since, as we have seen, they are by design perfectionist affairs whose central purpose is to generate athletic excellence. So even though these complex games are not ostensibly played with the appreciation of the fans in mind, they invite and welcome such watching and appreciation given their perfectionist character. That is why watching them is not only socially acceptable but socially encouraged, indeed even called for, unlike private affairs in which watching is not only unacceptable but socially verboten. As Stephen Mumford notes, this is in part what distinguishes sports from things like work and sex, in which watching is typically considered decidedly inappropriate.[31]

The fact that the relationship between players and spectators of sports is quite unlike that between producers and spectators of fictional works is no minor matter: it allows us to see and isolate one paramount feature of the actions that take place in sports that markedly distinguish them from anything that goes on in the fictional worlds that artists dabble in—indeed that would be ruinous of those fictional worlds. That feature is what Stear describes as the *authenticity* of game agency, by which he means those who play sports must "really" try to win, and those who watch them must be cognizant of this ineradicable fact about players' agency if they are to have any understanding, let alone a well-informed one, of what they are seeing. By contrast, he remarks further,

31 Mumford, *Watching Sport*, 132.

actors in theater and film, and authors of fictional literary works "typically dissemble," and the audiences they target must be cognizant of this ineradicable feature of fictional agency if they are to have any understanding, let alone a well-informed one, of what they are seeing or reading.[32] That it is essential in sports that players really try to win is of course crucial to my argument that it is the achievement value of winning and not make-believe that accounts for fans' enthrallment with game outcomes. For if trying to win were not an essential feature of sports agency, then the achievement value of winning would be null for both players and spectators alike. That is why, as Stear writes, "Where players feign effort . . . participant interest disappears or changes entirely."[33] So too, it goes without saying, does spectator interest and, to boot, appreciation. If dissembling of this sort in baseball and football were a legitimate feature of complex games like these, players would either turn to other venues to try to achieve difficult goals or would begin to think of themselves as primarily entertainers in the same way, say, as the Harlem Globetrotters do, interested only in putting on a good show rather than in excellently exercising their athletic capacities. And perhaps most importantly for my alternative take on winning, if fictionalists are right that it is spectators' make-believe that winning matters that explains their fixation on game outcomes, then authenticity would be "irrelevant to [their] ability to get behind a competitor."[34]

In fact, however, we know that enthusiastic fans do indeed regard the authenticity of gameplay as sacrosanct and, in turn, that nothing comparable to that authenticity is ordinarily at work in our enjoyment of traditional fictions.[35] That is because there is nothing that fans of this ilk detest more than when such

32 Stear, "Sport, Make-Believe, and Volatile Attitudes," 282. Just to be clear, that it is crucial in sports that players really try to win rather than feigning to try in this regard does not mean that athletic contests are allergic to any kind of dissembling. On the contrary, an important part of such competitions is that players try to disguise their actions by, for example, feigning to go in one direction all the while intending to go in another direction in order to throw off their opponents' efforts to stop them. Such deception and "trick" plays to mislead opponents are staples of all sports and are skills that are often pivotal to athletic success. The only dissembling that is anathema in sports is that concerning one's intent to give one's all to defeating the opposition.

33 Stear, "Sport, Make-Believe, and Volatile Attitudes," 282.

34 Stear, "Sport, Make-Believe, and Volatile Attitudes," 282.

35 Stear notes that sometimes actors are criticized for not giving their all to their portrayal of fictional characters, i.e., for "phoning in" their performances. But he insightfully points out that such cases of inauthentic performances differ from that of inauthentic gameplay. For the former, such inauthenticity has to do with their failure to convincingly convey the pretense called for rather than, as in the case of sports, not trying to win. As he writes, "it is one thing to criticize Sylvester Stallone for not playing Rocky 'believably', another for not trying to knock out Dolph Lundgren" ("Sport, Make-Believe, and Volatile Attitudes," 282).

dissembling occurs in sports. That explains why, Stear observes, fans detest match fixing. That is why they further typically detest anything and everything that puts in question the authenticity of the action on the field, and they detest it all the more if it has some significant effect on the outcome of the game. Take, for example, the phenomenon of what is called "flopping" in sports like football, which involves pretending that one has been fouled by one's opponent in order to draw a penalty from the referee. The writer John Lanchester comments on one such infamous example of this notorious tactic that incurred the immediate wrath of the spectators that took place in a quarterfinal football match between Italy and Belgium. An Italian player in that game, he writes, "collapsed with an apparently career-ending injury after being breathed on by a defender, only to bounce sheepishly back to his feet and trot off to join the celebration when Italy scored ten seconds later."[36] The spectators, he comments further, howled in derision at the player's resort to such fakery and bellowed in unison the choice word in their vernacular that they use to deride such misbehavior: "shithousery."

What such examples vividly show, I argue, is that enthusiastic fans do indeed care deeply that players try their level best to win and everything else it entails (among other things, that they play fairly). And the main reason why they care so deeply that players wholeheartedly strive to win and respond so viscerally to inauthentic gameplay, I argue further, is that they firmly believe that winning is indeed an intrinsically valuable achievement. Make-believe seems not only to have nothing to do with their palpable enthusiasm for such game outcomes but to be anathema to their appreciation of almost every facet of the game. That is why I am persuaded the fictionalist claim that spectators could not emotionally engage in sports and their outcomes with the great intensity that they do unless they imaginatively inflate the value of winning is intelligible, let alone forceful, only if one is already in the grip of that theory.

III

I now want to consider three objections to my argument that it is the intrinsic achievement value of winning that accounts for why enthusiastic fans obsess over game outcomes. The first objection comes directly from the pen of the fictionalist Baron-Schmitt, while the remaining two are, I surmise, objections that fictionalists would most likely press against my achievement view.

I begin with Baron-Schmitt's objection that argues that even if the achievement account of winning is correct, it does not provide a solution to the puzzling

36 Lanchester, "How Bad Can It Be?" 4. This is just one example of many other instances of inauthentic gameplay (too many to recount here) that are typically met with fierce resistance and criticism by those watching.

behavior of fans like his hypothetical fan Daniel, who intensely care about winning. Although Baron-Schmitt grants that achievement might explain why players care about winning, he maintains that it does not and cannot explain at all why spectators care about winning given the fact they are "mere" observers of games. For as mere observers of what goes on in games, they are unable to contribute in any direct or significant way to the actions that their teams take to defeat the opposition. As he writes, "Daniel does not accomplish anything if the Cowboys win. He just sits there, doing nothing to contribute."[37] So if winning matters because of its value as an achievement, he is persuaded fans would have "little or no reason to care about the outcomes of games, even if players do."[38]

The problem with this objection, however, is that it skirts the major question at issue here, which is why fans get so emotionally carried away by the on-the-field successes of the teams they follow. Rather than addressing this question head on, Baron-Schmitt instead poses the very different question of whether or not being a fan is itself a kind of achievement quite apart from what a fan's favorite players and teams achieve by winning the actual game itself, which is what fans so enthusiastically celebrate. That fans care about their teams' achievements even though they had nothing causally to do with those achievements is hardly surprising in itself, since we often care about other people's achievements outside of sports, even if we have had no hand in them. So while querying whether fans accomplish anything themselves if their teams win is an interesting question, it is not the question before us that I and, for that matter, my fictionalist counterparts are grappling with as to why fans are so taken with game outcomes. Therefore, as an objection to my own argument in favor of the achievement view to explain fans' preoccupation with game outcomes, it misses the mark.[39]

The second objection to my argument is that enthusiastic fans, especially partisan ones, follow the players or teams they do for entirely arbitrary reasons that have little to nothing to do with achievement as such. According to fictionalists, such fans are drawn to particular players or teams not by how they fare with regard to general qualities like the complexity of the challenges they encounter but rather by specific contingent factors: for instance, that their friends follow a particular player or team, that they fancy the uniforms of certain players or teams, that they root for players or teams that represent the particular

37 Baron-Schmitt, "Who Cares About Winning?" 250.

38 Baron-Schmitt, "Who Cares About Winning?" 250.

39 It goes without saying, I think, that the different question Baron-Schmitt raises here does not do any favor for the preferred fictionalist solution to the puzzle of sport. For if, as fictionalists argue, it is make-believe that accounts for why fans care about whether their team wins, then asking whether being a fan is an achievement in itself does not and cannot bolster their make-believe solution to fans' puzzling behavior because, again, it is not on point.

region of the country where they live, or countless other equally capricious considerations. Indeed, in many cases, being a sports fan, as Helfand suggests, is more like a heredity trait than it is a choice in that people are often fans of, say, the Philadelphia Eagles simply because their father grew up there, whereas "no one is a Bill Cosby fan because their father grew up in Philadelphia."[40] It is because there is evidently no logic to being a partisan fan that Walton claims, "you are not getting anything wrong if you root for the Tigers instead of the Blue Jays" (80). It could scarcely be otherwise, since for fictionalists, there is nothing more to fans rooting for the particular players or teams that they do than the arbitrary preferences that are the basis of fans' allegiance to them.

I do not think this objection is a forceful one for two reasons. For starters, it glosses over the already noted important fact that the great majority of enthusiastic fans follow only teams that have the requisite competence to achieve success in the relevant competition class to which they belong. To be sure, there are examples in which fans continue to follow teams despite their losing records, but they are clearly the exception rather than the rule.[41] In arguing that fans mainly follow teams that have what it takes to win, however, I am not denying they have other arbitrary motives for getting behind their favorite teams. Indeed, one thing this objection has going for it is that it calls attention to the fact that partisans often have mixed motives in this regard. But the one motive and reason I argue that almost all of them crucially share in common is that the teams they follow must be able to play the games well enough so that success is potentially at least within their grasp. That does not mean their attachment to teams rises or falls depending on the outcome of a single game but rather that it depends on the outcomes of a series of games over an extended period of time. Just how successful they need to be in this regard cannot be precisely quantified if only because we have to factor into the equation how successful their teams' competitors have been over the same period of time. Further, it

40 Helfand, "The Beautiful Mystery of Rooting for Aaron Rodgers," 11. That is why, he further opines, when Cosby went to jail, his fans quickly and seamlessly picked another comedian to follow; by contrast, those who grew up as Eagle fans typically remain Eagle fans even when, as is inevitably the case in sports, the team falls upon hard times.

41 And even some of these examples prove not to be the exceptions that they are often touted to be. Take, for example, the professional baseball team the Chicago Cubs, who are frequently mentioned in this light because for a time, they were known more for their failures than for their successes, but whose fans nevertheless stuck by them. As it turns out, however, the losing seasons they suffered through were just a small blip on their longstanding overall record, during which they achieved their fair share of success and even occasional dominance. In fact, to this very day, they "still hold the record for most wins by a major league club over any 10-year period" and are presently in contention for a playoff spot (Baker, *The New York Game*, 140).

goes without saying that given the highly competitive environment in which players of this exceptional caliber perform, no team, no matter how talented, is capable of dominating their opponents for very long. So long, however, as a team is able to maintain their competitiveness over the long haul, the best empirical evidence available to us strongly suggests that their fans will continue to support them, barring some egregious immoral or unjust offense on the part of the players and/or their major sponsors (athletic departments, professional sports owners, etc.).

The second reason why I do not find this objection persuasive is that there is nothing unusual or otherwise extraordinary about the fact that fans' initial attachment to teams is owed in part to arbitrary considerations rather than fully vetted rational ones. That is because many of our most valuable and cherished projects have similar arbitrary origins. For example, a common story that many successful and accomplished professionals, artists, university professors, and the like tell about themselves is how they started on their respective paths by sheer accident. It might have been, say, that they took a class in science or philosophy in college only because it happened to fit their schedules or because their friend did, but they subsequently fell in love with the subject matter and went on to pursue a career in medicine or university teaching themselves. The same narrative befits many sports fans who, as we have seen, might have begun to root for certain players and teams quite innocently because their parents did but subsequently fell in love with the team and began to follow them religiously. The moral of all these stories is clear: what begins with an arbitrary preference or desire to do or follow something more often than we might care to admit turns into our lifelong passions to play and/or listen to music or to engage in politics and/or write about it or to play sports and/or devotedly follow them. The fact that all of these life paths, which not incidentally bear all the marks of potential genuine achievements in their own right, can come to be by happenstance is no knock against them. Indeed, the idea that we care about the things that we do from the ground up on purely rational grounds is, as Harry Frankfurt avers, a "pan-rationalist fantasy."[42]

The third objection that fictionalists might plausibly raise against my account of the achievement value of winning concerns something that I have repeatedly emphasized as a cardinal feature of being an enthusiastic fan: such fans are nothing if not partial to the achievements and successes of players and teams they root for and typically indifferent to the achievements and successes of players and teams they do not root for. What fictionalists are likely to find problematic about such hard-to-miss partiality for certain players and teams is

42 Frankfurt, *The Reasons of Love*, 28.

that partisan fans are not ideal or even good fans but rather "epistemologically compromised" ones—unable and/or unwilling to value and appreciate the athletic achievements of the players they do not support. Since such engrained partiality is evidently, according to this criticism, a constitutive feature of being a fan of this kind, that means "believing without justification is just what it is to be" such a fan.[43]

Fictionalists might further argue in this same vein that if it is the achievement value of winning, as I suggest, that drives fans' over-the-top fixation on game outcomes, then it is impartial fans I should be singing the praises of, since they are open to and appreciative of whatever teams achieve the highest levels of athletic excellence. Instead of singling out partisan fans in this regard, I should single out two other types of fans that pass this impartial litmus test with flying colors—namely, so-called fair-weather fans and purist fans. Fair-weather fans, like partisan ones, root for and support particular players and teams, but unlike partisan fans, they root for and support only those players and teams that are presently playing the best. That is why they constantly shift their allegiances depending on who is winning at the moment. They are thus impartial in the sense that it is apparently athletic excellence alone that dictates their ever-shifting allegiance to particular players or teams. Loyalty and similar such features are simply not part of their normative vocabularies. However, purist fans do them one better on the impartial scale in that they are not even fans of particular players or teams who happen to be on a winning streak but rather fans of the sports themselves. As such, they differ from fair-weather fans in that they have no allegiance to any player or team but rather care only about the level of excellence achieved in the game itself. Purists are thus interested in athletic excellence *tout court*, which is why they "simply don't see the point of confining their sphere of care to one single club" that may after all not play well in the games they observe.[44] That is why, as Mumford argues, purists' perceptions and judgments of what is going on in a game and on its outcome are "far more trustworthy than that of partisans, which is, of course, why referees should always be neutrals."[45] That is further why Mumford thinks purists should eschew a competitive approach in favor of an aesthetic one in watching athletic contests, not caring about who wins but only about who plays the game in the most beautiful way possible. This gives purists, to his mind, a "gentler, less frenetic" appreciation of games in comparison to not only partisans but

43 Slater, "Epistemic Partialism in Sports Fandom," 7. Although Slater airs this line of attack against partisan fans in his essay, it should be noted that he ultimately rejects it.

44 Mumford, *Watching Sport*, 14.

45 Mumford, *Watching Sport*, 13.

fair-weather fans as well, which calls on what Hume calls our "calmer passions" but passions all the same.[46]

Now, there are two plausible ways to understand this objection. On the first, it is alleged that a partisan fan's partiality for the achievements of a certain team so skews their evaluative judgments of the achievements of the teams they do not favor that they cannot even begin to see their way to celebrating their successes. On the second, although a partisan fan's partiality for a certain team is not presumed to entail their complete indifference to or total lack of appreciation for the achievements of the teams they do not support, it is alleged that in light of the key role I assign to achievement in accounting for such fans' one-sided allegiance to particular teams, it is inexplicable why they are not instead impartial fans who are disposed to celebrate the achievements of *any* team. Suffice to say, if fictionalists can make either version of this objection stick, they expose an explanatory gap in the achievement view I peddle—one that can be effectively closed by swapping out my achievement theory for their make-believe theory. However, I do not think either version of this objection succeeds.

Let us start with the first version of this objection. The fictionalist gripe here, as noted, is that partisans' favoritism for certain teams runs so deep that it would be a wonder if they even noticed anything that opposing teams accomplish on the field, let alone applauded those athletic triumphs. But this criticism is exaggerated to say the least because it paints a false picture of partisans' actual beliefs and attitudes about the achievements of teams that are not their favorites.[47] It is of course incontrovertible that partisan fans want their teams to win. But it is a non sequitur to suppose that that disposes them to be indifferent to or to have false or distorted beliefs about the accomplishments of the teams they do not favor. We can see why this is so, curiously enough, by

46 Mumford, *Watching Sport*, 17.

47 It should also not go unremarked that the impartiality of fair-weather fans and purists is not as praiseworthy as it might at first blush appear. Though I do not have the space to go into greater detail here, there are at least two things I can and should say in this regard. First, fair-weather fans' fondness for winners may have less to do with their achievements and more to do with their own hedonistic mindset. That is because, as Glasgow writes, "there's more pleasure if your team always wins" ("The Real Fan," 7). By always being on the winning side, fair-weather fans conveniently manage to avoid all the heartbreaks and disappointments that partisan fans suffer when their teams lose. Second, while purist fans' aesthetic take on games is genuine enough, it is not at all clear that they can wholly eschew the partisans' competitive take on games that they decry. That is because what counts as a relevant aesthetic action on the field is one that is instrumental to the team's primary goal of winning. For example, a footballer's dazzling dribbling skills that lead to a successful shot on goal are rightly aesthetically lauded. By contrast, a footballer's dazzling display of dribbling skills that is superfluous to or even impedes her team's efforts to score a goal is rightly disparaged on both aesthetic and practical grounds.

considering how partisans view the gameplay of the teams they do support. For in this instance too, their obvious desire for their team to win does not cloud their critical judgment regarding their team's actual performance on the field. That is, it does not incline them to uncritically extol whatever actions their team execute during the game when such praise is unwarranted, to go easy on them when they play badly. On the contrary, partisans are acutely aware of and attuned to the limitations and weaknesses of their favorites. If anything, they tend to be hypercritical of them. That is why it is uncharacteristic for such ardent fans to hold back or soften their criticisms when their teams' play is less than stellar or to excuse any losses they incur because of their inferior play—by attributing them instead to bad luck or to biased and/or incompetent referees, or by making other such exculpating rationalizations. And it is because they are so critically disposed towards the teams they support that we should not be surprised that they are not in fact indifferent to or unappreciative of teams who defeat theirs. For when opposing teams outplay their favorites, partisans are given not only to criticizing the gameplay of their teams but also to grant, if only grudgingly, that the opposing teams were the better ones and deserved to win. As Kadlac avers, "the most resolute partisans among us can still appreciate a well-played game, even when their favorite teams lose."[48]

Furthermore, there is nothing about being a partisan fan that stands in the way of taking an interest in the athletic accomplishments of other teams and other players they heretofore had no interest in or even were unaware of. David Papineau's own experience as a sports fan is instructive in this regard. As an admitted partisan fan of several sports teams, he relates that his experience of following and admiring their athletic exploits is what sparked his interest in watching other sports he did not regularly follow, owing to the possibility that he might get to see other impressive athletic performances. This is, he tells us, what drew him to watch the one-hundred meter men's sprint finals in the 2008 Olympic Games even though he had no personal allegiance to any of the competitors. Nonetheless, the powerful running style of one of the runners, the justly famous Usain Bolt, quickly caught his attention, and he was overcome with joy when Bolt went on to win the event by a stunningly wide margin. Papineau's inability to take his eyes off Bolt and his great delight over Bolt's amazing victory were not because he followed Bolt—as noted, he did not—but rather, as he writes, "simply because excellence had prevailed."[49] And to allay any doubt as to either the tentativeness or soundness of his evaluative judgment of Bolt, he quickly adds, "any right-minded person would have felt the same."

48 Kadlac, *The Ethics of Sports Fandom*, 51.

49 Papineau, *Knowing the Score*, 99.

The second, less radical version of this fictionalist objection is not so easily disproved. In fact, my argument above that partisan fans are not only receptive to and appreciative of the achievements of at least some of the teams they do not support but even curious about and celebratory of the achievements of players and teams not previously on their radar makes this second criticism that much harder to disprove. For if, as I argue, achievement is key to explaining why partisan fans cheer as passionately as they do for the successes of their favorites and mourn their failures, then it is baffling why they do not respond in the same highly emotive way to the successes and failures of teams that they do not support, to include teams that outplay theirs.

This objection falls short, however, because there is in fact a very good reason why on my achievement account, partisan fans are partial to the successes of certain teams and not others. That reason has to do with the fact that no one becomes a dedicated partisan fan overnight. It is rather a process—to be sure, a longer one for some and a shorter one for others—but a process all the same. Remarkably, fictionalists have had little to say on this score, no doubt because they maintain the arbitrary preferences that attract people to teams in the first place cannot be rationalized but only fictionalized. That is why Walton maintains that the arbitrary preferences that lead people to "pick which teams or players to 'like'" come to seem to them actual and good reasons for picking them only by pretending they are such (76). But becoming a dedicated fan is not anything like the imaginative makeover that Walton envisions it to be. It is rather, again, a plausibly rational process that should be familiar to anyone who has ever become a fan of any worthwhile human undertaking. Walton is right, however, that people's initial attraction to a team, as noted and seconded in my response to the second objection above, often turns on something as random and parochial as "I watch the Dodgers because they play in the city in which I live." But from this inauspicious beginning, such an attraction then tends to develop over time (though of course not all those initially interested in a team continue to follow it), resulting in a special affinity for that team and undivided attention to the actions of the team's players during games and especially to how their performances competitively stack up against those of other teams. What draws sports fans closer and closer to their favorite players and teams at this stage are typically performance-related factors such as their style of play, their tenacity, their resilience when the chips are down, or their strategic ingenuity—all factors that attest to their team's capacity and resolve to excel on the field. From there, it is typically a short step to a fan's actual identification with the fortunes of that team, whereby it not only becomes the object of their intense focus and affection but an integral part of what Christine Korsgaard calls their practical identity—that is, an integral part of how fans see and think about

themselves and what they find and regard to be valuable about their lives.[50] At the culmination of this process of closely following their team's every move and getting more intimately acquainted with its players and how they perform as a collective unit, what were heretofore more or less arbitrary preferences for "picking" that team to watch morphs into what Stear rightly opines "look like *actual* [rather than inchoate or feigned] reasons for caring about [that] team and not another."[51] They look like actual reasons because they are in fact genuine reasons for caring more about the achievements of that team over others.

Stear goes on to argue that these "actual" reasons for caring more about one team over another are "comparable to reasons for preferring that one's own child win a prize over another's." And if quizzed why a parent has a reason to prize the achievements of their own child over another's, Stear insists, "'It's my child!' counts as good a reason as any."[52] He claims the same comparable answer—"It's my team!"—counts as good a reason as any for why partisan fans prize the achievements of their favorites over others. I think he is right about the first bit but wrong about the second. It is certainly true that a partisan fan's partiality for the achievements of a certain team and not others is analogous, *ceteris paribus*, to a parent's partiality for the achievements of their own child and not someone else's child. I think he is also right that it is not only perfectly acceptable and above board but as well entirely expected that a partisan fan roots for a particular team to outperform another, just as parents root for their own children to outperform others in some respectable pursuits. But where I split the difference with Stear is that while a parent's reason for prizing their own child's achievement over another's ("It's my child!") is no doubt as good as any other, a partisan fan's reason for prizing their own team's outperforming another ("It's my team!") is not as good as any other. That is because the indelible bond between a parent and child does not ultimately depend upon what my child accomplishes but simply on the fact "It's my child," full stop. That explains why most parents never truly stop loving and caring about their children even if those children never achieve anything or, worse, turn out to be huge disappointments. In stark contrast, however, a partisan fan's strong bond with their team does ultimately depend upon the team's competitive viability, on their favorite players' facility to prove their athletic mettle on the playing field. That is not to say such an enthusiastic fan's attachment to a team is not personal, but it is to say it is never *just* personal. It is also, as I argue, inescapably predicated on how *their own* team measures up to the opposition. Perhaps the clearest evidence, not to mention the textbook

50 Korsgaard, *The Sources of Normativity*, 101.

51 Stear, "Sport, Make-Believe, and Volatile Attitudes," 282.

52 Stear, "Sport, Make-Believe, and Volatile Attitudes," 282.

illustration, of the indissoluble relationship between a partisan fan's affective ties to their team and their estimate of their team's capacity for athletic excellence is the Florida Marlins baseball club, who in the span of one year (1997–98) went from a National League championship team with a large and passionate fan base to a "cellar dweller" with a rapidly shrinking and utterly demoralized fan base. The reason for their striking fall from grace was not a mystery, for their owner, Wayne Huizenga, sold off most of the members of the highly successful 1997 team to recoup financial losses that he had incurred over several years. Unsurprisingly, the 1998 Marlins compiled a dismal 54–108 win-loss record, and as noted, their fans abandoned the team in droves, leading to a precipitous decline in numbers for both their live games and television audiences.

Finally, I want to underscore something I commented on only in passing regarding what I describe as the rational process by which someone becomes a partisan fan of a certain team and, as a result, cares greatly about the team's achievements but not about those of other teams: the strikingly similarity with the rational process by which one becomes a devoted fan of and comes to care greatly about the achievements of a particular performer or performers but not those of others of some worthwhile endeavor. For just as people who, often on a whim, become initially interested in a baseball team like the Dodgers because they play in a city in which they live but then go on to become partisan fans of the team's athletic exploits, armed with actual reasons for rooting for them over, say, the Yankees, so the same goes for people who, often on a whim, go to a concert or an art gallery but are taken by what they hear or see and then become passionate fans of, say, the Beatles but not Bach or, say, of Picasso's abstract expressionist paintings but not Warhol's commercial or celebrity reproductions. If there is anything to be said for our plainly observable partiality for such widely different kinds of human achievements, it is surely that there is not something peculiar or rationally muddled or otherwise confounding about why some people end up being dedicated fans of the Dodgers but not of the Yankees, any more than that there is something rationally off or otherwise bewildering about why some people end up being fans of the Beatles but not of Bach, or of Picasso but not of Warhol. Lastly, if there is anything further to be said about why we humans care about some forms of human achievement but not others, it is surely that we finite humans are just not constitutionally equipped to care about everything worth caring about.

IV

To sum up, I argue that it is the intrinsic achievement value of winning that at bottom, explains why athletes pursue it so wholeheartedly and why enthusiastic

sports fans so tenaciously and passionately follow and support their efforts in this regard. For when all is said and done, sports are paradigm examples of human achievement that require the excellent exercise of our basic human capacities, especially of our bodily capacities and the remarkable things we are able to do and accomplish with our bodies in the carefully designed and demarcated spaces that sports provide for us. If I am right, then the fictionalist so-called puzzle of sports is in fact no puzzle at all and thus needs no solution, fictional or otherwise. Far from solving the puzzle of sports by providing a make-believe answer, fictionalists instead, I argue, merely beg the central question at issue here regarding just what players and fans see in sports and winning that explains why they take them so seriously and *play* and *attend* to them with such intensity and passion they seldom exhibit in their everyday lives.[53]

University of Southern California
wjmorgan@usc.edu

REFERENCES

Baker, Kevin. *The New York Game*. Alfred A. Knopf, 2024.
Baron-Schmitt, Nathaniel. "Who Cares About Winning?" *European Journal of Philosophy* 31, no. 1 (2023): 248–65.
Borge, Steffen. *The Philosophy of Football*. Routledge, 2019.
Bradford, Gwen. *Achievement*. Oxford University Press, 2015.
Frankfurt, Harry G. *The Reasons of Love*. Princeton University Press, 2004.
Gaut, Berys. *Art, Emotion and Ethics*. Oxford University Press, 2007.
Glasgow, Joshua. "The Real Fan: A Love Story." *Raven* 3 (2023): 1–21.
Helfland, Zack. "The Beautiful Mystery of Rooting for Aaron Rogers." *New Yorker*, September 7, 2024.
Hurka, Thomas. "Games and the Good." *Aristotelian Society Supplementary Volume* 80, no. 1 (2006): 217–35.
Kadlac, Adam. *The Ethics of Sports Fandom*. Routledge, 2022.
Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.
Lanchester, John. "How Bad Can It Be? John Lanchester on Cheating in Sport." *London Review of Books* 43, no. 15 (2021): 1–8.
Moore, Joseph G. "Do You Really Hate Tom Brady? Pretense and Emotion in

53 I want to acknowledge and thank the many helpful suggestions provided by the associate editor and the anonymous reviewers of this journal to a previous draft of this article.

Sport." *Journal of the Philosophy of Sport* 46, no. 2 (2019): 244–60.
Mumford, Stephen. *Watching Sport: Aesthetics, Ethics, and Emotion*. Routledge, 2012.
Papineau, David. *Knowing the Score: How Sports Can Teach Us About Philosophy (and What Philosophy Can Teach Us About Sport)*. Basic Books, 2017.
Searle, John. *Speech Acts*. Cambridge University Press, 1969.
Sher, George. *Beyond Neutrality: Perfectionism and Politics*. Cambridge University Press, 1997.
Slater, Joe. "Epistemic Partialism in Sports Fandom." In "Book Symposium: Alfred Archer and Jake Wojtowicz's *Why It's OK to Be a Sports Fan*." *Sport, Ethics, and Philosophy* (2024). https://doi.org/10.1080/17511321.2024.2367073.
Stear, Nils-Hennes. "Sport, Make-Believe, and Volatile Attitudes." *Journal of Aesthetics and Art Criticism* 75, no. 3 (2017): 275–88.
Suits, Bernard. "The Elements of Sport." In *Philosophical Inquiry in Sport*, edited by William J. Morgan and Klaus V. Meier. Human Kinetics, 1988.
von Kriegstein, Hasko. "On Being Difficult: Towards an Account of the Nature of Difficulty." *Philosophical Studies* 176, no. 1 (2019): 45–64.
Walton, Kendall. "'It's Only a Game!': Sport as Fiction." In *In Other Shoes: Music, Metaphor, Empathy, Existence*. Oxford University Press, 2015.
———. "Spelunking, Simulation, and Slime: On Being Moved by Fiction." In *In Other Shoes: Music, Metaphor, Empathy, Existence*. Oxford University Press, 2015.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.4431

POLITICAL FORGETTING

Alexander Bryan

> Clarendon counselled England's political leaders to "teach your neighbours and your friends … how to learn this excellent art of forgetfulness" and avoid finding new labels of division to ensure that "the old reproaches of Cavalier, and Roundhead, and Malignant, be committed to the grave."
>
> —Clare Jackson, *Devil-Land*

POLITICAL IDENTITIES and relationships are, among other things, matters of what and how we remember. The imagined communities of nation-states extend into the past, picking out events—battles and treaties, revolutions and reforms, atrocities and heroics—that inform what we take ourselves to be. So too for other identities—religious, racial, regional. Such practices of collective memory are taken by most (Nietzsche is a notable exception) to be both politically and ethically valuable—in some cases, even morally required.[1] Remembering injustices is essential for their remedy, and failing to remember may make us complicit in those injustices. More prosaically, having an understanding of those past events that have most shaped and continue to shape the political present might seem to be necessary for us to carry out our epistemic and political duties as citizens.

Sometimes, though, we are faced with appeals to *forget* certain aspects of our civic affairs. The quotation at the beginning of this article concerns one example, issued shortly after the monarchy was reestablished and Charles II crowned king of England—in other terms, in the aftermath of a long period of unprecedented political tumult preceded by a remarkably bloody conflict. The Earl of Clarendon was a veteran and a partisan of these conflicts, having served as a member of the Long Parliament and then as Charles II's political advisor. His call in 1660 to "learn this excellent art of forgetfulness"—and the accompanying Act of Free and Generall Pardon, Indemnity, and Oblivion, a broad ranging pardon "to bury all seeds of future discords and remembrance of the former, as well in [Charles II's] own breast as in the breasts of his subjects

1 Margalit, *The Ethics of Memory*, 9.

one towards another"—was, we can assume, issued with the aim of securing the new political settlement.[2] But we might also imagine it—and here we may depart from historical reality—as a sincere attempt to construct a postconflict political community on terms acceptable to all. Attempts to recast the categories of politics after conflict are common, and while they generally aim to achieve this by exposing rather than shrouding past injustices, some do incorporate notable injunctions to forget. To the case of the English Restoration we can add those of the Athenian Amnesty of 403 BC (in which the democrats issued a "decree [that] contained an explicit interdiction: it was forbidden to remember all the crimes and wrongdoing perpetrated during the immediately preceding period of civil strife," with all citizens required to take an oath not to do so) and the resolution to forget injustices and evils committed under the revolutionary regime in France issued by Louis XVIII upon his ascension to power in 1814.[3] Such cases of forgetting may also be rooted in action rather than exhortation, such as the toppling of statues upon the fall of a political regime, and may be implicit rather than directly commanded by political institutions or offices.[4]

In 1660, the categories of Cavalier, Roundhead, and Malignant would still have been highly relevant to public life, necessary for understanding ongoing political divisions and issues as well as the recent history of the polity. The injustices committed in this period would still have been fresh in the minds of those subject to them. Having these categories—as well the events and issues they mapped on to—in mind might appear not only to be required as a matter of civic engagement but as morally necessary. Political forgetting of the kind urged by Clarendon then seems to conflict with both our political and moral duties. This raises the question: Can such forgetting ever by justified?

I have two tasks in this article. The first is to provide a conceptualization of the kind of forgetting being urged by Clarendon—or, as I call it, *political forgetting*. I argue that we should conceive of political forgetting as a coordinated reduction in the salience of particular issues or categories in our civic affairs undertaken by the public in its capacity as the constituting people. I suggest that this is a more effective way of characterizing political forgetting than alternative models—namely, what I call the *belief-centered model* and the *prohibition model*. On my approach, political forgetting is undertaken by the constituting people for the purpose of reconciliation. This conception both accords with

2 Indemnity and Oblivion Act 1660, 12 Car. 2, c. 11, https://www.legislation.gov.uk/aep/Cha2/12/11/contents/enacted.

3 Connerton, "Seven Types of Forgetting," 61–62; and Ricoeur, *Memory, History, Forgetting*, 453–54.

4 Tanesini, "Collective Amnesia and Epistemic Injustice."

cognate discussions concerning public monuments and commemoration and shows that one source of the apparent conflict between political forgetting and our epistemic obligations as citizens is in fact illusory.[5]

The second task is altogether more ambitious. This is to outline the conditions under which political forgetting is justified. In brief, I suggest that these conditions are: (1) the transformation of the terms of political reality does not obscure terms or events that continue to structure society to the unjust disadvantage of any particular group; (2) that transformation is endorsed by all relevant social groups; and (3) the "forgotten" terms or events can be resurrected by agents in certain conditions. While demands made on others to forget are often opportunistic or cynical, I suggest there are at least in principle some circumstances in which political forgetting can be justified. As I show, these circumstances are rare. When it may be justified, political forgetting will usually not be undertaken in isolation but as one feature of a broader attempt to recast the terms of civic engagement. As even these cases will be quite rare, I suggest that the main practical benefit of clarifying the justificatory conditions of political forgetting is that it can help us diagnose cases in which unjustified calls to forget are made.

A final note before proceeding: throughout the argument, I refer back to the particular case of political forgetting urged by the Earl of Clarendon to illustrate different aspects of the view. I do so not to propose this as a paradigmatic case of political forgetting as I conceive it but purely for expositional purposes.

1. UNDERSTANDING POLITICAL FORGETTING

A starting point for developing a conception of political forgetting is to think about forgetting in our interpersonal lives. As Rima Basu has argued, forgetting is "an integral part of our moral duties to others."[6] We need to be able to forget some aspects of a person's history in order to have beliefs about them that reflect their current identity. Our personal identities are constituted through social practices of "holding" and "letting go," Basu claims (following Hilde Lindemann).[7] Treating others rightly involves letting go of, or forgetting, beliefs about them that conflict with their current identities.[8] We can think of

5 It does not, of course, cover all forms of forgetting. For instance, it does not incorporate cases in which documents are destroyed in order to prevent certain truths being made matters of historical record or public knowledge. My aim is to capture a particular kind of collective project of forgetting.

6 Basu, "The Importance of Forgetting," 472.

7 Lindemann, *Holding and Letting Go*.

8 Basu, "The Importance of Forgetting," 474.

occasions on which we do not do this as "holding failures": by "holding on" to certain information about a person's past we commit a doxastic wrong against that person. For instance, we plausibly wrong a trans person by referring to them by a name they no longer go by (i.e., deadnaming them); we might also wrong someone if we fail to forget information we learn about them by overhearing a private conversation.[9] We have reason to think both forgetting and failing to forget are morally evaluable, then; just as we wrong a close friend when we forget their birthday, we can also wrong them by *failing* to forget "those stories that no longer fit."[10]

An important feature of Basu's account of interpersonal forgetting is that it does not view forgetting as entirely involuntary. It is sometimes argued that our beliefs are not subject to our voluntary control. If this is so, the idea that we could ever have a duty to forget some event or concept would run up against the barrier of ought implying can. But as Basu argues, we do have at least some voluntary control over our beliefs, and the control we have is sufficient to ground moral evaluation of our beliefs.[11] If I really want to remember my friend's birthday, I can set reminders for myself or actively try to commit it to memory or check my recollection every so often. Similarly, I can and should take steps to avoid unjustly retaining information I should forget.

But we can also see that the kind of forgetting Basu argues is morally required is quite different to forgetting where you put your car keys or the postcode of the house you lived in as a child. We can in many cases fulfill our doxastic duties to others without *literally* forgetting—a feature of what I call the *belief-centered model of forgetting*. On such a view, forgetting can be achieved only when an agent no longer holds or has the capacity to access the relevant piece of information. The kind of forgetting we are morally required to practice in our interpersonal interactions need not involve this. Within some forms of forgetting, we can still have the capacity to recall that piece of information: what Basu calls *archival forgetting*, for instance, involves internally classifying some piece of information as unimportant or irrelevant to our purposes.[12] Doing this can make that information difficult to retrieve and prevent it from bearing on our normal practical reasoning. It thus seems to be an adequate way of classifying information like the deadname of a trans person. We may still be

9 Basu, "A Tale of Two Doctrines," 109–11, and "The Morality of Belief 1," 7. I thank an anonymous reviewer for suggesting the example of deadnaming. For discussion of a case similar to that of blamelessly overhearing, see Munch, "How Privacy Rights Engender Direct Doxastic Duties."

10 Basu, "The Importance of Forgetting," 474; and Lindemann, *Holding and Letting Go*, 45.

11 Basu, "The Importance of Forgetting," 487.

12 Basu, "The Importance of Forgetting," 475–76.

able to access the information again at some point in the future, and it would be inaccurate to say that we no longer hold that belief, but neither of these facts indicates that we are committing a doxastic wrong given the way in which we treat the relevant information.[13]

Like interpersonal forgetting, I suggest that a plausible account of political forgetting does not require the complete loss of access to the relevant categories or pieces of information. This would be deeply implausible given the continued salience and resonance of that information to individuals and society as a whole. It is difficult to imagine that any of Clarendon's audience could *really* be expected to actually forget the bloody civil war, the political chaos that preceded it, the unprecedented juncture of the interregnum, and the remarkable process of the reestablishment of the monarchy. Almost everyone in society would have known someone who died in the war, would have had their lives radically shaped by the political turmoil. Therefore, a plausible account of political forgetting must eschew the belief-centered model and mirror Basu's conception of interpersonal forgetting by incorporating forms such as archival forgetting—or, more plausibly in the case of the civil war, what Basu calls *siloing*, where we isolate and keep apart some memories from our practical reasoning because they are too painful.[14]

Relatedly, forgetting involves more than simply not raising an issue. One might imagine a group in which all members agree not to talk about some topic. We might call this the *prohibition model of forgetting*. Clearly, prohibition is a feature of some of the examples cited at the beginning of this article, with various levels of repression offered by different historical cases. In reality, Clarendon may well have been satisfied with this kind of forgetting. But we should reject this way of conceiving of political forgetting for two reasons. First, this appears less a model of forgetting and more a model of repression. While the absence of discussion of the relevant topic might on the face of it suggest forgetting, the topic might in fact remain just as prominent in the thinking of members of that society as it had been prior to prohibition. Second and relatedly, simply repressing talk of a topic is likely to be unstable, with that repression likely to falter when agents drop their guard or wish to provoke. But there is no reason to assume that collective forgetting is so unstable.

While drawing on interpersonal cases of forgetting can be useful, it is important to note a disanalogy in the goods promised by interpersonal and political forgetting. Interpersonal forgetting aims to bring about goods like

13 Bernecker, "On the Blameworthiness of Forgetting," 244.

14 Basu, "The Importance of Forgetting," 477.

privacy and forgiveness.[15] Neither of these is obviously the kind of thing that political forgetting could (or should) bring about. We cannot claim that our previous political commitments or actions are protected by our right to privacy—a right that covers our personal lives rather than our civic engagement. While forgiveness is sometimes raised in discussions of political transition from conditions of severe injustice or oppression, it would be highly implausible to think that *all* cases of political forgetting must aim this high.[16] Forgiveness is supererogatory in response to injustice, and appeals to forgive often trade on comprehensive ethical doctrines to which the state and citizens are not entitled to appeal when determining the terms of their cooperation.

This, though, does point us towards a better way of theorizing the good that political forgetting aims to bring about. This is not forgiveness but *reconciliation*.[17] Political forgetting should be understood as a collective endeavor aimed at changing the terms of cooperation in political life in the aftermath of serious injustice. Reconciliation is more usually sought through commemoration, truth commissions, and other processes focused on confronting or more fully revealing these injustices. These processes aim to unsettle established historical narratives constructed around severely unjust regimes, democratizing memory by exposing those aspects of that shared history that the powerful would have us forget. The notion that reconciliation could involve forgetting might seem to fly in the face of both our ethical judgments and the empirical evidence on the success of different ways in which societies have sought to reconcile after conflict or grave injustice. This is especially apparent in cases of atrocity, where forgetting might appear to be an act of complicity; a proper response to such cases involves sustained attention to the moral testimony of the victims and to the ways in which such crimes were permitted with the aim of ensuring no such atrocity can happen again. But as I outline in the next section, processes of forgetting may feature within these modes of reconciliation too. And just as in our interpersonal lives, forgetting might facilitate certain kinds of relationships, political forgetting may in some cases help create "living space for present projects."[18]

In the next section, I outline an account of political forgetting as a collective commitment to reduce the salience of some topics or concepts in our civic life. This is one aim of more standard features of reconciliation processes, which aim

15 Basu, "The Importance of Forgetting."

16 On the relationship between forgiveness and forgetting, see Margalit, *The Ethics of Memory*, ch. 6.

17 Brendese, *The Power of Memory in Democratic Politics*, 4.

18 Connerton, "Seven Types of Forgetting," 63.

to create the conditions for such reduced salience through mutual understanding, expressions of pain and dignity, and the confrontation of those perpetrators of injustice with the consequences of their actions. As I show below, the salience model of political forgetting does not mandate particular processes of reconciliation or suggest that forgetting can be justified if unaccompanied by other processes of confrontation and change.

2. A SALIENCE MODEL OF POLITICAL FORGETTING

In any society, political discussions and deliberations are informed by an understanding of what concepts, categories, and events are *salient*. That which is salient is worthy of or perhaps demands our attention.[19] As Susanna Siegel notes, the way information is presented in broadcast and print media follows a salience principle: to "make salient information that is important for the public to know about."[20] Our political discourse and everyday political reality are to a significant degree concerned just with those things that have been made or advanced as salient in this way. As I write this article, the most salient topic in European politics is the Trump administration's attempts to reshape the arrangements of the collective security of Europe. As the most prominent story in most newspapers and on most websites and television channels in recent days, this is presented as a matter to which the public should attend and that is *more* worthy of that attention than those matters that are given less prominence.[21] This organization of information by media outlets makes sense because it aims to enable citizens to direct their attention to those stories that most demand it and thus to direct the focus of public deliberation to those matters of greatest civic importance.[22]

There is an important sense in which we can think of salience in descriptive terms, as simply determined by the distribution of public attention. But it also has indispensable normative content. As the burgeoning philosophical literature on attention and salience makes clear, we can wrong people by

19 Siegel, "Salience Principles for Democracy."

20 Siegel, "Salience Principles for Democracy," 237.

21 If only temporarily. Of course, the salience attributed to a particular story need not reflect the overall significance of that event or issue relative to others.

22 There are of course other views on how media outlets should decide which matters they report on and which of those matters they grant prominence over others. Perhaps it is not importance but public interest that makes a matter newsworthy, or perhaps (as Siegel argues) the importance principle requires supplementation. I proceed with the importance version of the salience principle in mind, but the claims I make below should be accessible from a broad range of alternative positions.

misattributing the salience of particular characteristics.[23] Just as an individual makes a mistake when prioritizing (and thus as representing as more salient) their doctor's race or gender over their professional qualifications, so can a political community mistakenly or wrongfully attribute greater or less salience to some matter than is justified. The differences in the core political commitments of a polity, its institutional framework, and its historical development mean that what is properly viewed as salient across societies may vary. But within those constraints, there are some matters that we should as citizens prioritize and grant more attention to than others. Indeed, we must view salience in normative terms to understand why we should attend to some things rather than others *in our role as citizens*—things, that is, that bear on our ability to effectively satisfy the duties attendant to that role but that we would do not wrong in ignoring otherwise—and why we can reasonably be criticized for failing to recognize that something that is salient in another context is not relevant in political life generally or some specific area within it. The fact that some issue deserves more attention than some other does not mean that we are morally or politically required to grant it greater consideration in all circumstances, as there may be other reasons that give us cause not to.

Of course, politics is to a significant degree constituted by conflicts about which matters are most salient and how we should best conceive of them to reflect that salience. What is taken to be salient can not only change with events but be altered because we make an effort to pay attention to a different topic or feature of a situation, or because our values change.[24] We can disagree about what is worthy of attention as individuals and about what is deserving of our collective *civic* attention, but just as we individually can alter our attentional habits, so we can come to alter the focus of our joint attention. In most political communities, much disagreement begins from a position of agreement on the salience of other concepts or values and occurs against a background of the broader "social imaginary" of that society, to use Charles Taylor's term.[25] A social imaginary "incorporates a sense of the normal expectations we have of each other, the kind of common understanding that enables us to carry out the collective practices that make up our social life" and "some sense of how we all fit together in carrying out the common practice."[26] Social imaginaries can be altered as new narratives are asserted (often through social movements)

23 Munton, "Prejudice as the Misattribution of Salience"; and Whiteley, "Harmful Salience Perspectives" and "'A Woman First and a Philosopher Second.'"

24 Whiteley, "Harmful Salience Perspectives."

25 Taylor, *Modern Social Imaginaries*.

26 Taylor, *Modern Social Imaginaries*, 25.

and incorporated into that broader imaginary, coming "to be accessible to the participants in a way [they were not] before" and "eventually . . . to count as the taken-for-granted shape of things, too obvious to mention."[27] Despite the fact of political disagreement, then, political communities can alter what they pay attention to, as individuals do.

I think we can best understand political forgetting as an agreement to voluntarily reduce the salience of particular concepts or categories that are a part of our political life. This requires us to make an exception to the standard salience principle—that more important or significant information should be granted most salience—in pursuit of civic accord. The salience principle aims to guide our attention to those issues or properties that should be prioritized. Political forgetting involves an intentional collective departure from this principle with respect to some matter(s). This departure involves accepting a norm that alters the salience of the relevant topic in descriptive terms; it aims to move the matter down the order of priorities such that people do in fact view it as less salient and thus pay less attention to it. How strongly the matter is deprioritized is decided on a case-by-case basis by the relevant community (within some constraints, as I note below). In successful cases of political forgetting, this reduction in salience is constituted by a reduction in the attention paid to the matter by the citizenry. Reducing the salience attributed to that matter in social practices or institutions is instrumentally valuable in achieving this, as media organizations, public institutions, and political parties, among others, have the authority to assert to citizens the salience of some item and to constrain the extent to which they are able to reduce the attention they pay to it.

This way of thinking about forgetting nicely dovetails with some recent work on the role of monuments in our civic consciousness. This work largely begins from the premise that the historical processes by which public spaces come to be populated by monuments serve to "demarcate particular events, individuals and locations as especially significant to the nation's memory; and to materialise this in durable form."[28] The collective memory being constituted by these monuments might paradigmatically be national, but in many cases, it may instead (or also) be regional (as in the case of the erection of statues of Confederate generals throughout much of the antebellum American South in the twentieth century) or relate to some other mode of collective identity. The physical manifestation and visibility of monuments are an assertion of the *salience* of these individuals or events in that history, and as Michelle Moody-Adams among others argues, their presence "renders the past in ways

27 Taylor, *Modern Social Imaginaries*, 29; and Moody-Adams, *Making Space for Justice*, ch. 6.

28 MacDonald, *Memorylands*, 200.

that constructively shape the present."[29] While newspapers generally seek to advance the *immediate* salience of various items, public monuments and buildings make claims of *enduring* salience. Of course, a primary concern of both scholarly and public discussion of problematic monuments in recent years has been with promoting the salience of different kinds of considerations and perspectives with our collective histories and altering the public realm to better reflect this, as well as with challenging the honoring of morally objectionable individuals or individuals who are commemorated for morally objectionable reasons.[30] *Contra* some critics, these arguments and the public campaigns they discuss do not aim to erase or forget aspects of the past.[31] But note that some thinkers discuss forgetting as a motivation in the destruction or removal of monuments in the aftermath of revolution or regime change. As Alessandra Tanesini notes, using the example of the toppling of statues of Saddam Hussein in Iraq following the US-led invasion, removing such monuments can be seen as "an effective way of consigning [the regime] to oblivion."[32] This is plausibly an attempt not merely to alter the content of the collective history articulated by the built environment but also to reduce the salience of those figures and that regime within that history and spare citizens from seeing "objects that would cue memories of a past they wish to forget."[33] Sometimes, such action may be impermissible; the quick removal of symbols of the Nazi era after 1945 should, Anja Berninger argues, be thought of as a case of impermissible "defensive forgetting," as the removal of those objects was motivated at least in part by a desire not to fully reckon with or own up to that past.[34] Nevertheless, just as the physical environment can be altered by a community to promote or relegate the salience of particular concepts or information, so can our political concepts and reference points.

Any genuine recalibration of the salience of some topic—rather than, say, a straightforwardly political attempt to subdue the discussion of some historical events or salient categories—must be a collective affair. Without widespread buy-in, we are faced simply with a standard political conflict. Only widespread acceptance of the relegation of the salience of the relevant concept can support

29 Moody-Adams, *Making Space for Justice*, 138; and Schulz, "Must Rhodes Fall?"

30 Frowe, "The Duty to Remove Statues of Wrongdoers"; Lai, "Political Vandalism as Counter-Speech"; Lim and Lai, "Objectionable Commemorations," 3; and Schulz, "Must Rhodes Fall?"

31 Frowe, "The Duty to Remove Statues of Wrongdoers," 8–9; and Moody-Adams, *Making Space for Justice*, 144.

32 Tanesini, "Collective Amnesia and Epistemic Injustice," 214.

33 Tanesini, "Collective Amnesia and Epistemic Injustice," 195.

34 Berninger, "Erasing History as a Form of Defensive Forgetting," 548.

the coordination of such a shift. The agreement need not be universal to be effective. Clarendon's appeal to reduce the salience of the categories and divisions of the English Civil War and interregnum was aimed at the aristocratic ruling class, who held a monopoly on political power, rather than at the wider population. In modern democracies, we cannot expect all citizens to accept any democratic norms, or indeed that any particular topic is salient; it may instead come about through individuals each seeking to promote their own values or claims rather than coming to some prior mutual agreement.[35] Exactly the threshold required depends on the conditions of the society, but we can broadly say that the more widely distributed political influence is, the broader the agreement must be.

Finally, it is a crucial feature of genuine political forgetting that it is not a merely institutional affair but an act of *the constituting people*. Following the social contract tradition, we can distinguish between two senses of "the people": the constituting and the constituted people. One purpose of this distinction is to explain how the same referent of 'the people' can persist after radical constitutional change, such as the regime changes that took place at the start and end of the interregnum. The *constituted people*—that is, the people acting as a group agent within an institutional regime—is set aside and replaced in these cases. But, as Philip Pettit puts it, "the replacement of the constituted people need not return people to the state of nature … since the constituting people can retain enough coordination as a set of individuals to manage a smooth transition to a new constitution."[36] This coordination is the product of a "network of norms" that enables the constituting people to engage in one-off joint action without having a common will or an enduring organizational structure, and so may be quite minimal.[37] It can accordingly take a variety of different forms, with the possibilities for the joint action of the constituting people spanning the range of options for joint action in general; it may take a similar form to that of some individuals cooperating to carry a piano up some stairs or to the alternative form of those same people playing tennis, or it may emerge unplanned (as with a norm to greet others with a handshake rather than a hug).[38]

I think this distinction can illuminate two aspects of political forgetting: (1) how it can be undertaken; and (2) how it is consistent with the epistemic duties of citizens. To begin with 1, the distinction between the constituting and constituted people can explain the capacity in which citizens act when reducing

35 Pettit, *On the People's Terms*, 277–78.

36 Pettit, *On the People's Terms*, 291.

37 Pettit, *The State*, 206–11.

38 Pettit, *The State*, 209.

the salience of some matter. This might initially be puzzling, as individuals are acting neither in their institutional roles nor as entirely private actors. Because political forgetting is not a matter of formal institutional politics, it cannot (only) be undertaken by the constituted people.[39] But I suggest we can instead understand it as within the power of the citizenry acting as the constituting people. In this role, citizens may reorient the terms of democratic politics by altering the salience of particular items or concepts. On Pettit's characterization, the constituting people do not simply sit in the background ready to step in to respond to serious injustice but rather "jointly sustain the polity" by acting to populate civic and political offices and participating in the day-to-day life of democratic politics.[40] But there is also a long-term dimension of their role, which involves "subjecting government to the discipline of the considerations they valorize as relevant in public decision-making" (or, as Pettit puts it in *On the People's Terms*, "shaping government policies to the requirements of commonly endorsed norms").[41] Political forgetting is possible in these terms, as the constituting people might through a variety of methods relegate some considerations from the position of salience they once held. When the constituting people decides to alter the political salience of certain categories and concepts, they are deciding to subject the government to the discipline of some considerations over others, shifting the overall ordering of priorities. This extra-institutionally alters the terms of civic life, which, if successful, is then reflected in the institutional action of the constituted people.

The absence of a unified corporate will does not preclude this possibility. The acts of the constituting people are usually plural rather than the single act of an individual agent. These decisions are the subject of widespread acceptance—whether on the matter of which norms or considerations to which government policies should be subject or the matter of which proposed constitution should be adopted.[42] As the matter of which considerations should most strongly discipline public institutions is generally determined by competitive processes in

39 Of course, political institutions and office holders (like Clarendon) may encourage or attempt to shape a process of political forgetting. But while they can incentivize certain attitudes and aim to reproduce them, they cannot generate the kind of agreement that is required for a genuine process of political forgetting.

40 Pettit, "On the People's Terms," 694, and *The State*, 212.

41 Pettit, "On the People's Terms," 694, and *On the People's Terms*, 286.

42 Sieyès suggests that the constituting people might also act through the creation of "extraordinary representatives" not subject to ordinary constitutional rules or procedures. This body of representatives, operating as a "surrogate for an assembly of that nation" could indeed act as a corporate body while also acting in the guise of the constituting people. In the more ordinary course of things, however, the constituting people will indeed be understood as engaged in joint, plural action. See Sieyès, *Political Writings*, 139.

which citizens advance differing proposals, we can expect political forgetting to emerge either via the aggregated desires of many individuals to engage in such a process or (perhaps as part of such a process) through more explicit deliberation among citizens about some proposed course of forgetting.

Moving on to 2, I suggest the distinction between the constituting and constituted people can also provide a way out of an apparent conflict between processes of political forgetting and the epistemic duties we hold as citizens. Many scholars believe that citizens have epistemic duties of various kinds, which apply to our practices of voting, belief formation, protest, and deliberation. For instance, we owe it to our fellow citizens to take precautions to reduce the risk that we come to hold irrational or unjustified political views.[43] Duties of this kind appear to conflict with the practice of political forgetting, which may require us to ignore (or give less attention to) relevant information when forming our political views or advancing policies in public. While the epistemic duties held by most people in Restoration England would have been minimal, we can see this conflict more clearly by reimagining it as a democratic society in which all citizens hold some formal political power and basic political liberties. In this scenario, the kind of precautions citizens owe to each other to take would surely involve attending to the historical and political processes that explain current political issues, including those of the Civil War and interregnum. Political forgetting might, then, seem to be a violation of our responsibilities as citizens.

Understanding political forgetting as a task undertaken by the constituting people can help us to avoid this conclusion. Doing so might not initially seem to help much; after all, on Pettit's view, the constituting people are active "in the short haul of day-to-day, policy-by-policy decision-making," as well as in elections, political office, and social movements.[44] These "day-to-day" activities—like voting and standing in elections, or challenging the government in the courts, the media, or even through acts of civil disobedience—come with epistemic duties attached.[45] This demonstrates that the epistemic duties of citizens do not apply to citizens only in their role as the constituted people. But in their role as the constituting people, individuals may have quite different sets of epistemic duties or responsibilities, which apply to different aspects of that role. This is because, firstly, the constituting people have the prerogative to decide if and when there are reasons that justify departing from the usual salience principle and raising or reducing the attention we grant some consideration

43 Beerbohm, *In Our Name*, 151.

44 Pettit, *On the People's Terms*, 286.

45 Pettit, *The State*, 212–13.

based on the political values or aims of the community. Secondly, while citizens hold a duty only to provide public justifications for policies that appeal to considerations that can reasonably be expected to count as relevant grounds for all members of the community, this constraint does not preclude citizens from *refraining* from appealing to some consideration or from assigning it less significance than another.[46] As such, we can say that the epistemic requirements of citizenship do not rule out political forgetting, which remains an active option for the citizenry acting in its capacity as the constituting people.

Having outlined a model for conceiving of political forgetting, I now turn to the question of whether it can be justified and in what conditions.

3. Justified Political Forgetting

As noted in the introduction, there are strong *prima facie* reasons to be skeptical that political forgetting might ever be justified. There are two main grounds for this skepticism. The first of these concerns the distribution of burdens and benefits that emerges from a process of forgetting. While other modes of reconciliation appear to obviously promote the interests of those subject to injustice (by, for instance, enabling them to publicly express their experience of this injustice and have state institutions recognize the wrongs inflicted on them), reducing the salience of still-relevant political categories seems likely to disadvantage those who have suffered injustice. Suppressing categories of political significance can be a way of delegitimizing difference or criticism, or a means of generating a new common sense based on a partial historical narrative. For instance, in urging the English public to forget the categories of Roundhead and Cavalier, the Earl of Clarendon was, along with others, attempting to assert the reestablished monarchy's legitimacy by presenting it as continuous with the pre-interregnum constitutional regime. The methods used here of delegitimizing certain political actors and causes were clear, with the infliction of violence on the corpses of Oliver Cromwell and Henry Ireton just one "gruesome deterrent to future anti-monarchical agitators."[47] If processes of forgetting are always motivated by the desire of some social group to supress discussion of a topic or political opposition or to avoid confronting their role in some injustice, political forgetting will not be justifiable.

The second *prima facie* reason to be skeptical about the justifiability of political forgetting is that reducing the salience of important social categories might seem to license or even to mandate irrationality. Having a broadly accurate

46 Pettit, *On the People's Terms*, 253–54.

47 Jackson, *Devil-Land*, 378–79.

sense of the relative importance of different matters helps us to make better decisions individually and collectively. By jointly acting as if some matter were not as salient as it is, we deny ourselves relevant and important information.[48] We find this worry in the work of Paul Ricoeur, who, when considering cases of "commanded forgetting" such as that of the Athenian amnesty, asks, "Is it not a defect in this imaginary unity that it erases from the official memory the examples of crimes likely to protect the future from the errors of the past and, by depriving public opinion of the benefits of dissensus, of condemning competing memories to an unhealthy underground existence?"[49] Taking the Restoration again as a concrete example, Clarendon's call to forget would seem to have been asking people not to recognize the significance of the Civil War and interregnum and to ignore those aspects of monarchical rule that animated the "Good Old Cause" in the first place. Doing so would seem to inhibit the capacity of the polity to identify and perhaps resolve substantive and enduring political issues. Unless the topics in question actually come to be less relevant, then acting *as if* they are less relevant commits the political community to irrationality, falsity, and the reductions in welfare that may consequently accrue.

These are both important concerns, but while they certainly constrain the conditions under which political forgetting might justifiably be undertaken, they do not preclude its justification. I argue that political forgetting can in principle be justified when it meets certain conditions, which I suggest should assuage both of these concerns. These conditions are as follows:

1. The transformation of the terms of political reality must not obscure terms or events that will continue to structure society to the unjust and significant disadvantage of any particular group.
2. The transformation must be endorsed by all relevant social groups.
3. It must be recognized that it is legitimate for agents to raise the "forgotten" terms or events in public discussion in certain conditions.

The reasoning behind 1 is straightforward. The primary risk associated with political forgetting is that it will be used by the powerful—and in contexts where it is most likely to arise as an option, more precisely by the victors of a recent conflict—in order to recast political reality to their own partisan advantage by silencing discussion of a rightly salient topic. This would be unjust both as an epistemic and a political matter. As an epistemic matter, reducing the salience of a topic that continues to structure society economically and politically to a very significant degree would be to obstruct the capacity of those

48 Brendese, *The Power of Memory in Democratic Politics*, 13.

49 Ricoeur, *Memory, History, Forgetting*, 455.

disadvantaged by this structuring to understand their circumstances. This (and this is the political aspect of the worry) would also make it more difficult for this disadvantage to be challenged, as these matters may then become difficult to raise in public deliberation. Many practices of forgetting are unjustified for these reasons. Consider cases of what Alfred Archer calls *consigning to history*, which involves "the involuntary exclusion of certain people's identities from a community by narratives placing them in the community's past but not its present or future."[50] Examples of consigning to history include the development of narratives concerning portrayals of Indigenous people in North American popular culture and portrayals of industrial laborers in French narratives of deindustrialization. In both cases, these narratives present some identities—and those who continue to hold those identities—as inhabiting the past tense, rendering their place in the present invisible. These cases are unjustified, Archer suggests, because this invisibility "places them outside the temporal boundaries of this community."[51] Cases of consigning to history are wrong at least in part because the concepts, identities, and events they classify as belonging to the past and not the present remain relevant to the distribution of burdens and benefits. They make it harder for those affected to make their claims heard, thus violating 1. In very many cases, they are not endorsed by the group that is consigned to history and therefore also violate 2.

A question arises here: Do not all candidate case of political forgetting concern topics that have this kind of structuring effect? Topics or events that no longer effect political life require no special mechanisms to lose their salience. Political forgetting arises as an option only for issues that remain of high salience; it makes little sense for Clarendon to have urged his audience to forget events or political divisions that had already been forgotten or that were no longer relevant to public life. But 1 does not therefore rule out all political forgetting as unjustified. This is because not all topics of high salience continue to structure political reality in a way that is unjustly and significantly advantageous to some over others. They may not have an effect that significantly advantages some over others, or that advantage may be justified. Such concepts can continue to have high salience. In these situations, political forgetting remains potentially justifiable. But it is certainly true that this criterion rules as unjustified a very wide range of cases—including, plausibly, the forgetting urged by Clarendon, which seems very likely to have been to the undue disadvantage of antimonarchists. This should be regarded not as a defect of the account but as properly reflective

50 Archer, "Consigning to History," 7.

51 Archer, "Consigning to History," 7.

of the moral risks involved in political forgetting. It also points us to two specific modes of forgetting that more plausibly count as justifiable.

The first of these is forgetting that is undertaken as part of a broader program of change aimed at reducing the actual importance of the categories or concepts whose salience is to be reduced. A wide-ranging program of reconciliation that includes practices of truth telling, institutional reform, and the prosecution of perpetrators of severe injustice can include practices of forgetting that are premised on the operation of these other mechanisms. What distinguishes this from the ordinary application of the salience principle is that it is undertaken deliberately with an eye to altering the terms of civic engagement. The *prima facie* concerns noted at the beginning of this section are significantly abetted by marrying practices of forgetting with mechanisms aimed at reducing the importance of relevant categories. Indeed, if these mechanisms are successful, the kind of forgetting under discussion becomes rational.

One example of this is South Africa's postapartheid Truth and Reconciliation Commission (TRC). As P. J. Brendese has argued, "the TRC's proceedings reflect certain Nietzschean theoretical assumptions about how one remembers in order to forget."[52] The Nietzschean model of *willful forgetting* involves, in Brendese's terms, "the supplanting of an exclusive way of relating to the past."[53] The TRC undertook this first by democratizing memory, granting authority and validation to victims to publicly share their experiences. This disrupted the historical narrative that accompanied apartheid (in contrast to Clarendon's attempt to reinstate an "exclusive way of relating to the past"). But the amnesty to which these practices of truth telling was tied itself required a form of forgetting, or, as Brendese puts it, "a suspension of the remembrance of identities created by victimization."[54] Indeed, some of the criticism levelled at the TRC has been that it developed a narrative of reconciliation that failed to account for "memories resisting closure" and forgiveness.[55] What this example shows is that even processes of reconciliation that involve extensive processes of truth telling and public revelation of past crimes—that is, an actual reorientation of the salience of different concepts and topics in the shared life of the community—may have as one of their aims the kind of forgetting outlined above.

The second mode of forgetting is *prefigurative*, or undertaken as an experiment. In the absence of a broader program of change, a political community may embark on a project of political forgetting as an attempt to try something

52 Brendese, *The Power of Memory in Democratic Politics*, 43.

53 Brendese, *The Power of Memory in Democratic Politics*, 47.

54 Brendese, *The Power of Memory in Democratic Politics*, 50.

55 Brendese, *The Power of Memory in Democratic Politics*, 52.

new. Rather than definitively changing the terms of civic engagement, such a project could only justifiably be a trial run in which a citizenry explores the effects of reducing the salience of some concept or category. This kind of political forgetting is an act of collective faith, an invitation to other citizens to explore an alternative vision of the political community. This kind of forgetting *does* involve irrationality on the part of citizens, asking them to act as if some alternative state of affairs pertained. But it may still be justifiable given the mutual benefit that recasting the terms of civic life might generate.

Of course, both these forms of forgetting, especially the latter, are precarious and still run a risk of obscuring injustices impermissibly. Conditions 2 and 3 further specify the conditions under which political forgetting may be justified, and these further specifications constrain these risks.

Condition 2 requires that the reduction of the salience of the chosen topics or events is endorsed by the relevant social groups. This includes groups directly picked out by those concepts under discussion (e.g., Cavaliers and Roundheads in Clarendon's case), those whose identity and social position are a product of oppression strongly linked to those concepts (e.g., Black Americans in relation to the history of racial subjugation in the United States), and those who share a relevant feature with those groups (e.g., religious minorities in the aftermath of some schism between other religious groups).[56] There is a clearly a practical need for all relevant social groups to accept the reduced salience of some set of concepts for that reduction in salience to take place; if one group were to continue to operate as it did before, the salience of the topics would remain a matter of active political contestation, which is precisely the opposite of what political forgetting aims at. But the normative requirement involves a higher standard than this practical one. After all, groups can be brought to accept deeply unjust political settlements under threat of violence. The endorsement of groups must be given freely, and, though it need not be a matter of unanimous agreement among members, it should have wide-ranging support from members. The reduction in salience of some topic or concept must be understood by relevant social groups to be to their benefit (or at least not to their unjust disadvantage).

Finally, condition 3 requires that in certain conditions, the salience of the relevant concepts can be legitimately reasserted. A group that feels that the reduced salience of some concept is unjustly disadvantaging them is entitled to contest the agreement to relegate the salience of that concept. This is likely to itself be a matter of intense political contestation, but the right of communities

56 I leave open what precisely it means for an identity or social position to be a product of oppression or what might count as a relevant feature. The examples I provide indicate some of the cases I think are covered by any plausible answer to either of these questions, but the case I provide here does not rely on accepting any particular such answer.

to make such claims is a necessary component of any justified political forgetting. This does not mean that all such claims are warranted or acceptable to others, just that members of a society engaged in political forgetting—and especially those members of social groups referenced by categories being granted reduced salience—have the right to contest the justice of the new state of affairs. This is a necessary escape hatch to be used in case of unjustified or coerced cases of political forgetting. Threatening or inflicting outright violence on those who may contest the new state of affairs—as, for example, the reestablished monarchy did with the degradation of the corpses of Cromwell and Ireton—would certainly contravene this condition.

It is important to note here that condition 3 concerns the right to contest the reduced salience of a concept/topic and does not imply that political forgetting standardly involves such a steep reduction in salience that individuals are penalized for even mentioning that concept/topic. Any attempt to completely remove some concept/topic from political discourse will always violate condition *i* by virtue of severity; when a topic remains relevant to the understanding of an injustice, but citizens are not able to freely raise it without censure, those subject to that injustice are unduly disadvantaged. The possibility for contestation remains important even when the salience of a topic is only somewhat reduced, as that reduced salience can still have a serious effect on the terms of political discourse. If this effect is negative or unjust, it may be challenged.

Having outlined the conditions that apply to the justification of political forgetting, it is clear that there will be few real-life cases in which political forgetting is justified. These will most likely be cases in which forgetting appears as part of a broader package of changes to the terms of civic engagement and where the reduction of salience is moderate. But having a conceptual model of political forgetting and an account of the justificatory conditions that apply to such forgetting remains important. This is firstly because in these cases, justified forgetting may come with significant benefits. Reducing the salience of some concept or topic may, in tightly circumscribed situations, productively reconfigure the terms of political engagement within a society. But the account of political forgetting offered here also has a second useful function that is more broadly applicable. It can be used to evaluate real-world charges to forget—in particular, to demonstrate why they should be rejected. When faced with calls to forget, we can use the account to spot when and why those calls might be unjustified. Perhaps the proposed forgetting would be to the disadvantage of some particular group or is not endorsed by that group or makes it overly onerous for that group to raise concerns about some topic or using some concept in the future. Whatever the fault may be, the account sketched above helps us to identify what exactly is wrong with such proposals.

4. CONCLUSION

I have argued that we should conceive of political forgetting as a coordinated reduction in the salience of some topic or concept undertaken by the public in their role as the constituting people and for a set of justificatory conditions according to which we can evaluate proposals to forget. The model I have outlined is highly demanding; many calls to forget aspects of our political life neither fit the conceptual framework nor meet the justificatory conditions. This should not be regarded as a defect of either aspect of the model or as a reason to dismiss the concept of political forgetting entirely. Rather, it should be seen as a reflection of the high stakes involved in such cases and of the frequent unjustified calls to forget that are a feature of political life. My account vindicates the judgment that these calls are unjustified. But it leaves open the possibility that some forms of political forgetting may be justified. It also suggests that those cases will likely involve more than just a reduction in salience, encompassing instead a broad array of measures designed to alter in fact as well as thought the terms of civic engagement within a political society.

More generally, the account of political forgetting I have offered presents questions of what and how societies should relate to their pasts as not just thoroughly political but as requiring a particular kind of democratic politics.[57] Civic memory is always a matter of political struggle and conflict, but civic forgetting of the kind I have been concerned with here operates at one step of removal from these everyday practices of political contestation—informed but not determined by them. Acts of political forgetting are one extraordinary means that political communities can take to reorient the terms of their shared life, to create the possibility that their institutions will step beyond the reenactment of past conflicts and create a shared democratic future.[58]

University of Cambridge
ajb387@cam.ac.uk

57 Brendese, *The Power of Memory in Democratic Politics*.

58 I am grateful to two anonymous reviewers and an associate editor from this journal for their helpful comments. I am also grateful to the Isaac Newton Trust, which provided funding (grant number NTACDF5) for the fellowship I have held while developing this research.

REFERENCES

Archer, Alfred. "Consigning to History." *Philosophers' Imprint* 24, no. 19 (2024): 1–14.

Basu, Rima. "The Importance of Forgetting." *Episteme* 19, no. 4 (2022): 471–90.

———. "The Morality of Belief 1: How Beliefs Wrong." *Philosophy Compass* 18, no. 7 (2023): e12934.

———. "A Tale of Two Doctrines: Moral Encroachment and Doxastic Wronging." In *Applied Epistemology*, edited by Jennifer Lackey. Oxford University Press, 2021.

Beerbohm, Eric. *In Our Name: The Ethics of Democracy*. Princeton University Press, 2012.

Bernecker, Sven. "On the Blameworthiness of Forgetting." In *New Directions in the Philosophy of Memory*, edited by Kourken Michaelian, Dorothea Debus, and Denis Perrin. Routledge, 2018.

Berninger, Anja. "Erasing History as a Form of Defensive Forgetting." *Journal of Applied Philosophy* 42, no. 2 (2025): 547–63.

Brendese, P. J. *The Power of Memory in Democratic Politics*. Boydell and Brewer, 2014.

Connerton, Paul. "Seven Types of Forgetting." *Memory Studies* 1, no. 1 (2008): 59–71.

Frowe, Helen. "The Duty to Remove Statues of Wrongdoers." *Journal of Practical Ethics* 7, no. 3 (2019): 1–31.

Jackson, Clare. *Devil-Land: England Under Siege, 1588–1688*. Penguin, 2022.

Lai, Ten-Herng. "Political Vandalism as Counter-Speech: A Defense of Defacing and Destroying Tainted Monuments." *European Journal of Philosophy* 28, no. 3 (2020): 602–16.

Lim, Chong-Ming, and Ten-Herng Lai. "Objectionable Commemorations: Ethical and Political Issues." *Philosophy Compass* 19, no. 2 (2024): e12963.

Lindemann, Hilde. *Holding and Letting Go: The Social Practice of Personal Identities*. Oxford University Press, 2016.

MacDonald, Sharon. *Memorylands: Heritage and Identity in Europe Today*. Routledge, 2013.

Margalit, Avishai. *The Ethics of Memory*. Harvard University Press, 2002.

Moody-Adams, Michelle. *Making Space for Justice: Social Movements, Collective Imagination, and Political Hope*. Columbia University Press, 2022.

Munch, Lauritz Aastrup. "How Privacy Rights Engender Direct Doxastic Duties." *Journal of Value Inquiry* 56, no. 4 (2022): 547–62.

Munton, Jessie. "Prejudice as the Misattribution of Salience." *Analytic Philosophy* 64, no. 1 (2023): 1–19.

Pettit, Philip. "On the People's Terms: A Reply to Five Critiques." *Critical Review of International Social and Political Philosophy* 18, no. 6 (2015): 687–96.

———. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge University Press, 2012.

———. *The State*. Princeton University Press, 2023.

Ricoeur, Paul. *Memory, History, Forgetting*. University of Chicago Press, 2004.

Schulz, Johannes. "Must Rhodes Fall? The Significance of Commemoration in the Struggle for Relations of Respect." *Journal of Political Philosophy* 27, no. 2 (2019): 166–86.

Siegel, Susanna. "Salience Principles for Democracy." In *Salience: A Philosophical Enquiry*, edited by Sophie Archer. Routledge, 2022.

Sieyès, Emmanuel Joseph. *Political Writings: Including the Debate Between Sieyès and Tom Paine in 1791*, edited by Michael Sonenscher. Hackett, 2003.

Tanesini, Alessandra. "Collective Amnesia and Epistemic Injustice." In *Socially Extended Epistemology*, edited by J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard. Oxford University Press, 2018.

Taylor, Charles. *Modern Social Imaginaries*. Duke University Press, 2003.

Whiteley, Ella Kate. "Harmful Salience Perspectives." In *Salience: A Philosophical Enquiry*, edited by Sophie Archer. Routledge, 2022.

———. "'A Woman First and a Philosopher Second': Relative Attentional Surplus on the Wrong Property." *Ethics* 133, no. 4 (2023): 497–528.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.3548

STEREOTYPE THREAT AND ALIENATION

Ana Barandalla

IN THIS ARTICLE, I expand our understanding of the normative weight of stereotype threat and point where next to further that inquiry. My focus is the *realization* of the threat of stereotype threat, a phenomenon that I call *succumbing to stereotype threat*. I show that in a common type of stereotype threat, succumbing to stereotype threat is characterized by a corrupt relation of the individual to herself. That baneful relation is, I argue, what stereotype threat is a threat of. We also learn that that relation is facilitated by the individual's social environment, in which the individual is an active participant. The individual, then, contributes to the forces that set her against herself. That social dimension is the area that stands in need of further investigation. The analytic resources that help me to that point are not, on their own, well equipped to explore that social terrain, but they do suggest an apt tool for it: the notion of *alienation* as understood by Karl Marx. I argue that the corrupt relation of the individual to herself at the heart of succumbing to stereotype threat, complete with the role that the individual's social environment—including the individual's own input—plays in that relation, is captured in the notion of alienation. That notion is part of a tradition well honed in the ways in which one's social environment mediates one's relation to oneself; hence, it is suitably placed to examine the social dimension of stereotype threat. The notion of alienation, then, lends a name to the wrong of succumbing to stereotype threat and thus to the threat of stereotype threat, and it has the credentials to further enhance our understanding of those phenomena.

Stereotype threat is the threat you experience at the prospect of confirming a negative stereotype that applies to you and that is widely held in your community.[1] The threatening prospect might be that of confirming the stereotype to yourself, to others, or both.[2] The experience of stereotype threat is characterized by a range of cognitive and emotional upheavals, including fear and

1 Steele, *Whistling Vivaldi*, 120–21; Nguyen and Ryan, "Does Stereotype Threat Affect Test Performance of Minorities and Women?" 1314–34. Steele's book is the *locus classicus* on stereotype threat for lay readers.

2 Shapiro et al., "Stereotype Threat," 87–105. See also Blum, "The Too Minimal Political, Moral, and Civic Dimension of Claude Steele's 'Stereotype Threat' Paradigm."

anxiety, preoccupation to defeat the stereotype, pressure on your self-worth and sense of self, decreased self-confidence, and self-doubt.[3] The outcome is often a performance well below your best.[4]

Stereotype threat was first identified by psychologists, notably Claude Steele and colleagues, but its normative import has attracted the interest of philosophers.[5] The bulk of philosophers' concern has been the consequences of stereotype threat, such as the role of stereotype threat in shaping the demographics in different spheres of social life, not least professional philosophy.[6] More recently, some authors have called for attention to be directed to additional normative strands. Ron Mallon emphasizes the importance of discussing stereotype threat in terms that resonate with individuals' own experiences of stereotype threat.[7] Stacy Goguen urges us to acknowledge the ways in which stereotype threat might compromise one's own sense of worth as a human being.[8] And Lawrence Blum advocates conceiving of stereotype threat in ways that acknowledge the social structures in which stereotypes are maintained.[9]

My own interest lies in yet another normative aspect of stereotype threat, but it is an offshoot of my study that Mallon's, Goguen's, and Blum's requests go some way to being met. My interest is in the inner normative structure of stereotype threat itself, quite apart from its consequences. I take on a specific and commonplace type of stereotype threat—one comprising cases in which (1) the threat pertains to the prospect of confirming the stereotype at least to yourself; and (2) the stereotype at play asserts that your epistemic competence is poor—that people belonging to your group are simple or stupid. These two features combine to form a first-person iteration of Sally Haslanger's "epistemic objectification," and so we might say that the threat in those cases of stereotype threat is the threat of epistemic *self*-objectification.[10] My specific concern is the

3 Steele, *Whistling Vivaldi*, 118–21, 124.

4 Steele, *Whistling Vivaldi*, 124.

5 Steele, *Whistling Vivaldi*.

6 Alfano, "Stereotype Threat and Intellectual Virtue"; Antony, "Different Voices or Perfect Storm"; Beebee, "Women and Deviance in Philosophy"; McKinnon, "Stereotype Threat and Attributional Ambiguity for Trans Women"; Saul, "Implicit Bias, Stereotype Threat, and Women in Philosophy"; and Schouten, "The Stereotype Threat Hypothesis."

7 Mallon, "Stereotype Threat and Persons."

8 Goguen, "Stereotype Threat, Epistemic Injustice, and Rationality" and "Expanding Our Picture of Stereotype Threat." Goguen, in particular, is interested in some of the same issues I pursue here.

9 Blum, "The Too Minimal Political, Moral, and Civic Dimension of Claude Steele's 'Stereotype Threat' Paradigm."

10 Haslanger, "Objectivity, Epistemic Objectification, and Oppression, 280."

realization of that threat—being overpowered by stereotype threat and coming to accept that the stereotype is true of you.[11] I call that phenomenon *succumbing to stereotype threat*. So if in Steele's terms, stereotype threat involves "arguing against the stereotype," succumbing to stereotype threat is the phenomenon of losing that argument, of being defeated by the stereotype and thus coming to accept that in virtue of belonging to the group to which the stereotype is attached, you do indeed possess the negative traits assigned to you.[12]

Viewed this way, succumbing to stereotype threat is the *formal* threat of stereotype threat. To arrive at a *substantive* picture of the threat of stereotype threat, I seek to make sense of the dense normative scene there presented. Because I treat succumbing to stereotype threat as fundamentally an epistemic phenomenon, I look to accounts of epistemic normativity for help. I find that none of the mainstream accounts—the norm truth of belief, the rationality norm of belief, virtue epistemology—can give us the understanding we seek because none can produce an integrated picture of the phenomenon's normative features. A constructivist account, in contrast, can. It produces a picture in which succumbing to stereotype threat is an instance of a perverse relation of the individual to herself, set in an environment in which the individual is herself participant. This relation, with those features, is, I argue, an instance of alienation as understood by Marx. At least in the commonplace type of cases under consideration, then, the substantive threat of stereotype threat is alienation.

My discussion makes some strides towards answering Goguen's, Blum's, and Mallon's calls: it lays bare the way in which stereotype threat can wreck one's sense of self; it pays heed to the complicated relationship between the individual and the forces that maintain the stereotypes that assail her; and because it recognizes the phenomenology of succumbing to stereotype threat as integral to it, much of the description of the phenomenon is couched in terms recognizable to the individual. But my discussion also adds to Blum's plead in particular. Our foray into the social dimension of succumbing to stereotype threat opens up a broad view of the social structures that maintain stereotypes and of the role of individuals, including stereotyped individuals, in those structures. It is clear that no full understanding of the normative profile of stereotype threat is possible without a finer-grain grasp of that setting. I end this article with a proposal that the notion of alienation is apposite to lead us into that further research. I begin with a fuller presentation both of stereotype threat and of succumbing to stereotype threat.

11 I use 'accept' in a nonfactive sense.

12 Steele, *Whistling Vivaldi*, 123.

I

The conditions in which stereotype threat occurs vary to some extent, but the archetypal setting is this: you are performing or are about to perform a testing activity; you are aware of belonging to a social group stereotyped as being bad at that activity; performing well in that activity is important to you.[13] It helps our discussion to have an example at hand. Suppose that your society draws a stark line between those born in March and those born in any other month of the year. Having a birthday in March or not is considered one of an individual's most important features, and from cradle to grave, people's lives are profoundly shaped by whether they are thought to have been born in March or not. Those born in March are known as "Marchers." There is a well-developed, well-entrenched stereotype attached to Marchers. It includes the notion that, unlike non-Marchers, Marchers are not good thinkers. They are thought to be weak creatures who at the slightest cognitive strain, collapse into a heap of emotions and cannot string a thought together.

You are a Marcher and are about to give a lecture. Doing a good job of it is important to you: you are presenting a cherished project, and setting your ideas forth publicly is indispensable for establishing yourself in your profession. As you are introduced to your audience, your being a Marcher is variously salient: you are spoken to and about in a manner reserved for Marchers, you display the demeanor expected of Marchers, and you are groomed and garbed as standards for Marchers demand. You look, behave, and are treated like a Marcher and are acutely aware of it. With it, the idea that Marchers are thought not to be good thinkers looms large in your mind and, you expect, in your audience's. As these thoughts gain shape, pressure in your self-confidence builds up. Soon, cracks begin to appear: What if you have been kidding yourself about your capacity to do this and are now about to expose yourself as a phony, with all these eyes on you? A surge of anxiety attends, and the besieging thoughts gather potency: Maybe there is a good reason why Marchers are thought not to be made for this kind of thing? Your blood pressure and heart rate rise sharply, making it harder to stay focused and composed.[14] Instead of readying yourself for your talk—rehearsing your main points, owning your space, acknowledging your audience—you strain every sinew to defeat the thought that the stereotype might be true and to assuage the creeping anxiety.[15] You are experiencing stereotype threat.

13 Steele, *Whistling Vivaldi*, 126; and Nguyen and Ryan, "Does Stereotype Threat Affect Test Performance of Minorities and Women?" 1315.

14 Steele, *Whistling Vivaldi*, 119, 121.

15 Steele, *Whistling Vivaldi*, 111, 123, 126.

II

The situation might then unfold broadly in one of three different ways. In the first, you somehow succeed in defeating the stereotype: you regain your confidence, establish a good rapport with your audience, and do great. In the second, you wrestle with the stereotype for the duration of your talk: you neither get on top of it nor are quite beaten by it. Your performance is hampered, but you are not left thinking that you are a failure. In the third, you succumb to stereotype threat: you are crushed by the force of stereotype threat and resign yourself to the thought that you cannot do this because you are a Marcher. This last is the outcome that stereotype threat is geared to bringing about. Let us see how.

The specific threat in our example is that, in virtue of being a Marcher, you are not a good thinker, that you are intellectually weak and easily overcome by emotion. Now, in suffering from stereotype threat, you are experiencing strong emotions—fear and anxiety—and they are interfering with your thinking. *We* know that these are manifestations of stereotype threat, but they are also what the stereotype anticipates. Of course, the stereotype anticipates those upheavals as part and parcel of your being a Marcher, while under stereotype threat, you experience them regardless of whether the category of Marchers even exists in its presumed ontological footing. But your experience of those disruptions is not sensitive to their provenance. And even if you are schooled about the phenomenon of stereotype threat, and you are also aware that you are currently under its spell, you might well remain only too sensitive to the possibility that these symptoms are overdetermined, that they are manifestations both of stereotype threat and of your being a Marcher.[16]

Furthermore, because you are eager to defeat the stereotype, your sensitivity to pertinent evidence heightens.[17] You look around and find your environment saturated with evidence in alignment with the stereotype: by far the majority of your audience are non-Marchers, as are most of the authors whose work you engage with; the walls of the hosting room are decorated with portraits of non-Marchers—their non-Marchness and sage gazes all of a piece; the room itself is named after a non-Marcher, and so is the building that houses it. The thinking prowess of non-Marchers is everywhere hailed and displayed, whereas, if only by dint of omission, the opposite is promulgated about Marchers.

In the midst of all that, support for the stereotype piles up, and the possibility of defeating it recedes. Intensely alive to that fact, your anxiety takes another jump, further interfering with your cognitive activity and consuming

16 Saul, "Implicit Bias, Stereotype Threat, and Women in Philosophy," 46–47.

17 Steele, *Whistling Vivaldi*, 124.

its resources. Your performance is suffering badly now; you are stumbling upon your thoughts and your delivery. Things are increasingly turning out just as the stereotype predicts, and your resistance is ever more futile. There comes a point when the evidence seems overwhelming, and there is only one reasonable conclusion: the stereotype is right—you are a lousy thinker and an emotional mess.[18] You lose the argument. You succumb to stereotype threat.

The force of stereotype threat, then, is epistemically self-fulfilling: it threatens you with the prospect that the stereotype is right, and it primes you to collect all and any evidence for it while generating a steady supply of that evidence. In that light, succumbing to stereotype threat is the natural terminus of stereotype threat. And the struggle in stereotype threat—the struggle against the prospect that the stereotype is true—is a struggle against succumbing to stereotype threat. In this sense, then, succumbing to stereotype threat is the *formal* threat of stereotype threat.

Let us now turn to analyzing the phenomenon of succumbing to stereotype threat. This tells us what the substantive threat of stereotype threat is, as well as what it amounts to.

III

Succumbing to stereotype threat is marked by the acceptance of a belief—in our case, the belief that in virtue of being a Marcher, your epistemic competence is poor. For ease, let us refer to it as the *core belief* of succumbing to stereotype threat. Given the pivotal role of this belief, I approach succumbing to stereotype threat as an epistemic phenomenon. And because our concern is normative, to help make sense of that phenomenon, I look to accounts of epistemic normativity.

In addition to the core belief, succumbing to stereotype threat exhibits a range of other normative features. One of them is your own role in it. The process leading up to and of finally succumbing to stereotype threat involves agential input if any epistemic activity does. You *guide* your deliberation this way and that, you *try* to refute the stereotype, you *persist* in that effort, you finally *give in*—you *accept* the proposition that the stereotype is right and your epistemic competence is poor. Another normative feature is a sense—vague for now—that you are somehow a victim, that in succumbing to stereotype threat, you are in some way wronged or that a harm has been done to you that you neither deserve nor consent to. Finally, there is the phenomenology of

18 Steele, *Whistling Vivaldi*, 125–26.

succumbing to stereotype threat.[19] This has not been part of my exposition, so let me fill that gap now. To do so, I draw both on psychological studies and on the relation between stereotype threat and succumbing to stereotype threat.

Social psychologist Gordon Allport finds that coming to accept negative stereotypes about oneself provokes "low self-esteem, low expectations, low motivation, self-doubt, and the like."[20] Victoria Valian, another psychologist, remarks that "being told that the group to which we belong is deficient in some area … is a threat to self-esteem."[21] Succumbing to stereotype threat involves both those things: through the stereotype of Marchers, your society tells you that the group to which you belong is deficient in epistemic competence; the stereotype of Marchers is therefore a negative stereotype, and it is one that, in succumbing to stereotype threat, you come to accept. We would expect, then, that succumbing to stereotype threat is accompanied by low self-esteem, low expectations, low motivation, and self-doubt.[22]

Reflection on the relation between stereotype threat and succumbing to stereotype threat delivers a similar view. We have seen that the lead-up to succumbing to stereotype threat just is the struggle that is the hallmark of stereotype threat. We have also seen that this struggle is driven by your feeling threatened at the prospect of the stereotype being correct: you are fearful and anxious about it because your self-value is at stake. This is why stereotype threat bites harder the more you care.[23] Given that in succumbing to stereotype threat you take the object of those fears and anxieties to be confirmed, we would expect the feelings of stereotype threat to crystalize into feelings of dejection, low self-esteem, and self-doubt—just as the psychologists tell us.

The nature of those ills—particularly low self-esteem—deserves attention. Anna Bortolan, following Matthew Ratcliffe, explicates self-esteem as an all-enveloping sense of one's place in one's environment, a sense that underpins one's meaningful engagement with the world.[24] *Lack* of self-esteem, then, must involve a sense of *not* having a place in the world, of your interactions with the world

19 Throughout, what I mean by 'phenomenology' and its cognates is the character of experience rather than the philosophical discipline. My thanks to an anonymous referee for encouraging me to make this explicit.

20 Quoted in Steele, *Whistling Vivaldi*, 46.

21 Valian, *Why So Slow?* 152.

22 The claims in this paragraph allow for the possibility of exceptions. Where that is the case, the unfolding discussion would not apply. My thanks to an anonymous referee for pointing this out.

23 Steele, *Whistling Vivaldi*, 126.

24 Bortolan, "Self-Esteem and Ethics." See Ratcliffe, "Existential Feeling and Narrative."

being devoid of meaning.[25] If we assume that integral to your sense of self is a sense of how you relate to your environment and that the exercise of your agency is a matter of engaging with the world in ways that are meaningful to you, Bortolan's analysis indicates that lack of self-esteem flattens your sense of self and thus diminishes your agency.[26]

The phenomenon of succumbing to stereotype threat, then, presents a rich normative terrain: it features a core belief that bears your agential input, yet you are also a victim, and it burdens you with a crippling phenomenology. The attempt to make sense of succumbing to stereotype threat is an attempt to better understand that terrain, to illuminate those normative features, and to see why they should cluster together. To this end, let us summon the mainstream accounts of epistemic normativity, beginning with the truth norm of belief.

IV

The truth norm of belief maintains that the fundamental epistemic norm is that beliefs be true.[27] Accordingly, the pressing normative question concerning the phenomenon of succumbing to stereotype threat is: Is it or is it not the case that in virtue of being a Marcher, your epistemic competence is poor? I am going to stipulate that the answer is that it is not.[28] On that basis, the truth norm of belief tells us that the fundamental wrong of succumbing to stereotype threat is that the core belief is false. From this it follows that since you are agentially involved in coming to accept the core belief, the wrong in succumbing to stereotype threat is *your* wrong, for you have brought it about. And that too is what you are a victim of: your own epistemic turpitude.

As regards the phenomenology of succumbing to stereotype threat, the truth norm of belief does not have much to say. That is because the phenomenology of succumbing to stereotype threat arises in a relation between the individual and her beliefs, while the main concern of the truth norm of belief is the relation between the content of a belief and its truth conditions—a relation that is phenomenologically inert. The phenomenology of succumbing to stereotype threat, then, must be regarded as a psychological matter, not an epistemological one. Any epistemological interest must be limited to whether it helps or hinders meeting the norm of truth going forward. But this is a matter

25 Bortolan, "Self-Esteem and Ethics," 6.

26 Bagnoli, "Moral Constructivism"; and Velleman, *How We Get Along*.

27 McHugh, "The Truth Norm of Belief."

28 It would be harsh to regard this stipulation as creating a paradox: one false belief a poor epistemic agent doest not make. See also Blum, "The Too Minimal Political, Moral, and Civic Dimension of Claude Steele's 'Stereotype Threat' Paradigm," 153.

of the consequences of the phenomenology of succumbing to stereotype threat, so not germane to our brief, which is confined to the nature of succumbing to stereotype threat as part of the inner normative structure of stereotype threat.

I conclude that the truth norm of belief does a poor job of making sense of the phenomenon of succumbing to stereotype threat. It overlooks the phenomenology of succumbing to stereotype threat, and it passes a troubling indictment on you, rendering you blamable for succumbing to stereotype threat and thus for the situation in which you are a victim. To be sure, this account might well elicit our agreement on finding the falsehood of the core belief an important normative feature. But we might shrink from regarding it as the fundamental wrong of succumbing to stereotype threat. Were we to witness someone succumb to stereotype threat, I doubt our main concern for her would be that she has acquired a false belief. I expect we would think there is something more at play. But this account of epistemic normativity is devoid of resources to tell us what.

V

Let us next look to the reason norm of belief. This norm commands that beliefs stand in rational relations to other beliefs, regardless of their truth value.[29] On this view, the foremost normative question about succumbing to stereotype threat is whether the core belief stands in a rational relation to the beliefs upon which it is based.

As presented in our example, it does. You pick up facts that you take to have a bearing on the truth of the stereotype, you strive for consistency, and finally you accept the fateful belief precisely because of its rational pull—because that is where the weight of the evidence and of the relevant considerations fall. Granted that much of what you take as evidence for the truth of the stereotype are in fact manifestations of stereotype threat. But we can have no rational quarrel with that because as far as your experience of those manifestations of stereotype threat is concerned, and hence as far as your epistemic grasp of them goes, they are indistinguishable from manifestations of the stereotype. And the contextual considerations and additional evidence lean heavily on interpreting those manifestations of stereotype threat as manifestations of the stereotype. Under the rationality norm of belief, then, the core belief of succumbing to stereotype threat is unblemished. There is nothing fundamentally wrong in succumbing to stereotype threat.

This renders your agential involvement as nothing but praiseworthy. No danger of blaming the victim here. Indeed, since this framework does not pick

29 See, e.g., Zangwill, "The Normativity of the Mental."

up any wrong, the idea that you are somehow a victim seems confused. The phenomenology of succumbing to stereotype threat too is puzzling. Not that we would learn much about it if it were not thus seen anyway. Since the rationality norm of belief is concerned with relations between the content of different beliefs rather than between beliefs and the individual whose beliefs they are, it too would be silent on why succumbing to stereotype threat should have the phenomenology it has. Once again, that phenomenology is seen as a psychological issue, and any epistemological interest rests on its consequences (e.g., whether it reduces your ability to subsequently strike a rational balance between beliefs), but that, again, strays from our remit—namely, finding out how and why the normative features of succumbing to stereotype threat fit together.

I conclude that the rationality norm of belief is ill equipped to deliver the understanding of succumbing to stereotype threat we seek. We might sympathize with its verdict that in succumbing to stereotype threat (in accepting the core belief), you do nothing wrong, that your conduct there is justified. But the overall assessment that nothing fundamental is wrong in succumbing to stereotype threat surely jars. Furthermore, rather than illuminating any sense in which in succumbing to stereotype threat you might be said to be a victim, this view renders it a mystery, as it does the phenomenology of succumbing to stereotype threat, about which it would be quiet anyway. Let us move on.

VI

Next up is virtue epistemology. Virtue epistemology focuses on the relation between an agent, on the one hand, and her beliefs and general epistemic activity, on the other. Here, the fundamental normative epistemic standard is the expression of intellectual virtues.[30] Beliefs and epistemic activity generally are right to the extent that they express epistemic virtues and wrong to the extent that they express epistemic vices. As for what counts as epistemic virtue, virtue epistemology is divided into two contrasting camps: *reliabilists* and *responsibilists*. Reliabilists hold that the central characteristic of epistemic virtues is their connection to truth, while responsibilists maintain that epistemic virtues are distinctive for being constitutive, or expressive, of epistemic flourishing. Both camps characterize epistemic vices as epistemic attitudes that inhibit the virtues. For both reliabilists and responsibilists, typical examples of epistemic virtues include open-mindedness, fairness, humility, and attentiveness. Examples of epistemic vices include stubbornness, arrogance, and narrow-mindedness.

30 Battaly, "Virtue Epistemology," 640.

In this framework, the key question for us is whether your coming to accept the core belief of succumbing to stereotype threat is expressive of epistemic virtues or of epistemic vices. On the face of it, it looks like the virtues are in charge. You initially respond to the dreaded proposition with resistance—that is, the proposition that the stereotype is right and your epistemic competence is poor. But that resistance is born out of judiciousness, not obstinacy: it would be reckless not to demand high levels of certainty for a proposition that carries such dire implications. Throughout, you are resolute in following the evidence, you diligently take it on board, and when you deem it incontestable, you have the courage and humility to admit defeat and accept the conclusion, even as it comes as a body blow.

On the strength of that description, we might be tempted to rate your coming to accept the core belief as a model exercise of intellectual virtues. However, that appraisal would satisfy neither reliablists nor responsibilists. Reliabilists would point out that you arrive not at a true belief but at a false one. Responsibilists would decry that far from epistemic flourishing, succumbing to stereotype threat presents us with a constellation of pathologies, most notably those comprised in its phenomenology. Your coming to accept the core belief therefore cannot be approved by either camp of virtue epistemology. Presumably, their verdict would be that even if what we see in your accepting the core belief is a virtuous process, it is not virtuous *enough*. In this case, the fundamental wrong of succumbing to stereotype threat is wanting in epistemic virtue.

This returns us to some of the issues we encountered with the truth norm of belief. Since you are agentially involved in coming to accept the core belief, any failure to exercise the virtues is *your* failure. The fundamental wrong of succumbing to stereotype threat is therefore *your* wrong. And, again, you are a victim of your own virtue deficiency, and you are to blame for it.

As for the phenomenology of succumbing to stereotype threat, virtue epistemology might seem to have the edge over the truth and the rationality norms of belief. This is due to two key tenets held by responsibilists in particular. The first is that you are responsible for your epistemic virtues.[31] The second is that epistemic virtues are profoundly important for the individual: they are "properties of the soul"—that is, properties that are "most deeply constitutive of herself," thus "vitally connected with a person's identity."[32] Suppose you believe in those two tenets. Suppose too that you regard your epistemic competence as reflective of your epistemic virtues. Accepting the core belief is tantamount to accepting that you have failed to develop your virtues and thus that you have

31 Zagzebski, *Virtues of the Mind*, 125.

32 Zagzebski, *Virtues of the Mind*, 104, 135, 125.

failed to cultivate deep and important aspects of yourself. This kind of self-directed blame about something so critical as the kind of person you are might well carry the phenomenology we observe in succumbing to stereotype threat: low expectations, low motivation, self-doubt, low self-esteem.

However, this attempt to account for the phenomenology of succumbing to stereotype threat levels yet another charge at you. The core belief makes a claim about your nature in being a Marcher. That is not a claim about you in any sense for which you are responsible. You cannot do anything about being a Marcher, and you cannot do anything about the nature of Marchers. Any self-directed blame about your epistemic competence in virtue of your being a Marcher is therefore misplaced. This would be denounced by both reliabilists and responsibilists: the former would point out that the belief that you are responsible for your epistemic status is false, the latter that the engendered phenomenology is far from an instance of flourishing.

All in all, virtue epistemology still falls short of providing a unified account of the phenomenon of succumbing to stereotype threat. It rules—improbably, I suggest—that your succumbing to stereotype threat is a manifestation of insufficient virtue on your part and that you are a victim of that virtue deficiency; it can explain the phenomenology of succumbing to stereotype threat only by imputing a further failing to you; and you are to blame for it all.

VII

I next consider constructivism. There is not a full-fledged constructivist account of epistemic normativity per se, so I shall develop one as we go along, drawing mostly on Christine Korsgaard's work. On constructivism, normativity arises in a relation in which the individual stands to herself. As such, this relation sets the standards of correctness for what you do. On Korsgaard's constructivism, that relation is one of authority.[33] Thus, what you do is right or wrong depending on whether it is an exercise of your authority to yourself or not. The critical question for us, then, is: Do you, in coming to accept the core belief and thereby succumbing to stereotype threat, exercise your own authority to yourself, or not?

Before we can address that question, we need to see what your exercise of your authority to yourself looks like; and before we can do that, we need a clearer picture of the sense in which you are an authority to yourself. So let us start there. You are an authority to yourself in the sense that only you can decide what to do. This springs from the dichotomous structure of the self, comprising a passive part and an active part. Very roughly, the passive part is where

33 Korsgaard, *The Sources of Normativity* and *Self-Constitution*.

phenomena are visited upon you, while the active part encompasses what you do—that is, it encompasses the phenomena with which you identify.[34] The passive part presents you with impulses to do or to believe something or other; the active part decides whether to do as per those impulses.[35]

For example, a scene in front of you might prompt you to believe that a tree has been felled; or, another example, you may find yourself pining for a chocolate truffle in the kitchen. Upon becoming aware of the impulse to believe that a tree has been felled or of the impulse to have the chocolate truffle, you face the question of whether to do as per those impulses or not: whether to settle on the belief that the tree has been felled, whether to have the truffle. Let us call this *the question of what to do*. On this view, the posing of the question of what to do is something with which you identify; it is something where you are active. This question could not be answered by the passive part of you. Any answer proposed by the passive part would be met with a corresponding question of whether to do as per *that* proposal.[36] Only what counts as you, only where you are active, can answer the question of what to do.[37]

According to Korsgaard, the necessity that your question of what to do be answered only by you *is just what normativity is*.[38] Any other normative commitment, any other normative claim on you, is but an elaboration of this relation. That renders that need—the need that you answer your question of what to do—as the fundamental requirement to which you are subject. It thus sets the standards of correctness for what you do: whatever you do, your answer to the question of what to do must be authored by yourself. That, then, is the sense in which you are an authority to yourself. Let us now see what the exercise of that authority looks like—what it is for you to be the author of the answer to your question of what to do.[39]

We have seen that what counts as *you* is that with which you identify, the active part of you. Korsgaard tells us both (1) that the active part of you is reflection and (2) that you identify with what you find "expressive, or representative, of yourself"—that is, what reflects your values, including their ordering.[40] (Henceforth, when I talk about values, I include their ordering, unless

34 Korsgaard, "The Activity of Reason," 23.

35 Korsgaard, *The Sources of Normativity*, 93.

36 Korsgaard, *The Sources of Normativity*, 93.

37 Korsgaard, *Self-Constitution*, 72, and *The Sources of Normativity*, 100.

38 Korsgaard, *Self-Constitution*, 2–4, and *The Sources of Normativity*, 103–4.

39 This is a compressed interpretation of Korsgaard's discussion about practical identities. See Korsgaard, *The Sources of Normativity*, 100–107, 236, *Self-Constitution*, 20–26, and "The Activity of Reason," 23.

40 Korsgaard, *Self-Constitution*, 75, and *The Sources of Normativity*, 100.

otherwise stated.) Taken together, those statements tell us that you are the author of the answer to the question of what to do when you reach that answer through a deliberative process that reflects your values.

The exercise of your authority to yourself, then, depends on what your values are; and what your values are is a contingent matter but for one: your value of yourself as the author of the answer to the question of what to do. On Korsgaard's account, that value is both necessary and fundamental.[41] It is *necessary* for your agency because it is the value that sets you off in deliberation. And it is *fundamental* because it is what bestows your other, contingent values their practical status: you are required to reflect your values in your answers to the question of what to do *because* you are required to be the author of that answer. Whatever contingent values you have, your answer to the question of what to do counts as expressing your authority to yourself only if it reflects the primacy of your self-value as the author of the answer to that question—in other words, only if it reflects the primacy of your authority to yourself.

This picture invites a reformulation of our central question. Initially we asked whether in coming to accept the core belief and thus succumbing to stereotype threat, you express your authority to yourself. We can now see that that amounts to asking whether in coming to accept the core belief and thus succumbing to stereotype threat, you reflect the supreme value of your authority to yourself.

To address that question, we need a fuller picture of your self-value as the author of the answer to the *epistemic* question of what to do—the question of what to believe. Its roots, according to Korsgaard, are in your self-value as the author of the answer to the *practical* question of what to do.[42] Let us begin there. In conceiving of yourself as the author of the answer to your practical question of what to do, you take yourself to be able to guide your practical engagement with the world by that answer.[43] That requires that your conception of the world be unified—that it present the various things in the world and their relations in a way that you can trace and establish, particularly their logical, spatiotemporal, and causal relations.[44] For example, if you decide to cross the road, you must presuppose a grasp of how things like the breadth of the road, the width and pace of your gait, the speed of the oncoming traffic, relate to each other.

41 Cf. Street, "Coming to Terms with Contingency."

42 Korsgaard, "The Activity of Reason."

43 Korsgaard, "The Activity of Reason," 34. See also Taylor, "Rationality."

44 Korsgaard, "The Activity of Reason," 34.

But this conception of the world does not come to you already formed. As we saw earlier, you experience impulses to believe this or that, and you then have to decide whether or not to do as those impulses prompt you to. Korsgaard sees this space between the potential ground for a belief and the belief itself as one that "presents us with both the possibility and the necessity of exerting a kind of control over our beliefs. . . . We can actively participate in giving shape . . . to the conception of the world in light of which we act."[45] In other words, when we ask the epistemic variant of the question of what to do, we set ourselves the task of producing a conception of the world "in light of which we can act," and a conception of the world in light of which we can act is a unified conception of the world.

Viewed that way, the function of epistemic deliberation is to build a unified conception of the world. And so when presented with the impulse to believe that *p*, your decision of whether to believe that *p* or not must come down to whether your picture of the things in the world and their relations makes more sense to you—i.e., is more unified—with the belief that *p* than without it.[46] Finding that unity is a matter of striking a reflective epistemic equilibrium amongst the relevant items of your epistemic economy.[47] Those items include, crucially, though not exclusively, the belief that *p* or its absence or negation, the considerations that lead you to countenance whether *p* in the first place, your preexisting beliefs, and any new beliefs you garner along the way.

Your self-conception as the author of the answer to the epistemic question of what to do is, then, more specifically a self-conception as the author of your unified conception of the world. Therefore, you exercise your authority to yourself in your answer to the epistemic question of what to do when that answer manifests the primacy of your self-value as the author of your unified conception of the world.

A familiar type of example serves to illustrate that claim. Suppose you have been informed by the police that your daughter's fingerprints have been found all over a crime scene, and she alone has access to the site. They put it to you that she is guilty, but you resist that idea. In one version of this story, you are sure

45 Korsgaard, "The Activity of Reason," 32.

46 'Without the belief that *p*' is meant to encompass both the explicit rejection of the proposition in question (as when one believes that it is not the case that January is a rainy month in Baghdad) and the absence of a belief—either positive or negative—about that proposition altogether (as when one's conception the world does not include a belief about whether January is a rainy month in Baghdad).

47 See Rawls, *A Theory of Justice*. Rawls uses the notion of *reflective equilibrium* in relation to moral judgments, but it describes aptly the give and take involved in the acceptance or rejection of a proposition too.

that you heard your daughter's footsteps in her bedroom at around the time of the crime, and you remember that she left her handbag (where the keys to the site would have been) on the train a couple of weeks ago. These thoughts are not in equilibrium with the proposition that your daughter is guilty. Your stance is that for as long as that lack of balance remains, you withhold acceptance of the proposition that your daughter is guilty. In another version of the story, you have always passionately and defiantly maintained that your daughter is an ace young woman, a testament to your excellent parenting skills, and you will be damned if a bunch of uniformed nobodies were to challenge that. You *dismiss* the information conveyed to you by the police and reject the belief that your daughter is guilty.[48]

In the first version, you reach your conclusion in pursuit of a unified conception of the world, while in the second, you do it in pursuit of certain comforts. In the first version, your deliberation reflects your values and their ordering; you identify with it. In the second version, it does not, and so you cannot identify with it, not fully anyway.[49] That, then, is what it looks like to produce an answer to the epistemic question of what to do that reflects the supremacy of your self-value as the author of your unified conception of the world. We are now ready to look at whether in coming to accept the core belief of succumbing to stereotype threat, you do reflect that value.

VIII

When looking to see whether in coming to accept the core belief you might reflect the supreme value of yourself as the author of your unified conception of the world, we look to see whether you might come to accept the core belief on the basis that doing so best produces a unified conception of the world.

We have seen that your pursuit of unity in your conception of the world involves striving for equilibrium in your epistemic economy, with the most salient elements including: the proposition in question (or its absence or negation); the considerations that led you to entertain it as a possible proposition to believe; your preexisting beliefs; and the new beliefs you beget along the way. Among your preexisting beliefs, some exert greater gravitational pull than others. Some of the factors contributing to that extra pull might be spurious

48 Bagnoli uses a similar example to discuss self-deception, where the agent is also understood to be an authority to herself ("Self-Deception and Agential Authority").

49 This of course touches upon the question of the extent to which a decision must reflect your values for it to count as yours and the extent to which it must not for it to count as a wrong decision. I cannot enter this debate here. For Korsgaard's own position, see *Self-Constitution*, ch. 8.

(comfort, seduction), but others might be lawful—that is, they might spring from your commitment to author a unified conception of the world. Among the possible lawful factors are (1) how reliable you take the sources of the given beliefs to be and (2) the degree of integration of those beliefs within your epistemic landscape—that is, how much support those beliefs lend to and receive from your conception of the world at large.

Stereotypes with the ubiquity and significance of the stereotype of Marchers score high on those two fronts. Those who introduce, reiterate, and enforce those stereotypes to you are bound to include people of authority to you, such as your parents and wider family, their social milieu, your teachers, and persons of high social standing. Almost by definition, you will have regarded these sources as reliable. And those stereotypes are reflected explicitly and implicitly in local mores, received wisdom, hermeneutic resources, social practices, cultural expressions, and so on, all of which come to form and furnish your epistemic landscape just in the process of growing up and becoming a member of your community. By the time you reach epistemic maturity, those stereotypes are well rooted in your conception of the world, supporting and being supported by extensive networks of beliefs and ways of going about in your surroundings. Stereotypes such as that of Marchers, then, come with a layer of credibility that might lawfully wield considerable force when calibrating reflective epistemic equilibrium.

Consider too the beliefs that you collect as you strive for that equilibrium. Many of them ensue from the epistemically self-fulfilling tendency of stereotype threat (see section 1 above)—beliefs resulting from your being primed to taking evidence generated by stereotype threat as evidence for the truth of the stereotype. These beliefs, we can now see, are not only fueled by stereotype threat; they fit hand in glove with the stereotypes that populate your existing epistemic landscape. That, in turn, reinforces those stereotypes, thus increasing their weight in the attainment of any epistemic equilibrium.

All in all, your epistemic economy is configured towards accepting propositions that aver the truth of the stereotype. And so it is perfectly possible that the idea comprised in the core belief—that in being a Marcher your epistemic competence is poor—better contributes to a unified conception of the world than its absence or negation. If it is on this basis that you come to accept the core belief, then you succumb to stereotype threat in a manner that reflects the preeminence of your self-value as the author of the answer to the epistemic question of what to do. Since that is how you exercise of your authority to yourself, and exercising your authority to yourself is the standard of correctness for what you do, it follows that you might succumb to stereotype threat while doing the right thing.

IX

That is not to say that on the constructivist picture, there is nothing wrong in succumbing to stereotype threat. We have seen that what renders your acceptance of the core belief right is that that acceptance reflects the primacy of your self-value as the author of the answer to the question of what to do. But the proposition that you are epistemically incompetent implies that you are *not* good at providing the answer to the question of what to do. The content of the core belief, then, undermines the belief's own authority. In other words, your coming to believe that your epistemic competence is poor undermines the authority behind your coming to believe that proposition. Coming to accept the core belief and hence succumbing to stereotype threat, then, sets you up in a paradoxical relation to yourself: you employ your authority to yourself to undermine that very authority. On our constructivist framework, that is the fundamental problem with succumbing to stereotype threat.

That paradox is unlike familiar ones such as "Don't believe anything I say" or "Always do the opposite of what I say." If I uttered those instructions to you, you would be trapped in a paradox for only as long as you were trying to decide how to do what I ask you to. But you can always "cheat" your way out of it by simply walking away. If doing what I ask you to do involves not doing what I ask you to do, you can jolly well decide to do neither. Life has other things to offer. You can just move on. But in the case of the core belief, you cannot walk away because that belief concerns your capacity to think and thus to decide at all, including to decide to walk away. Any decision to break free from the conclusion that you are not good at reaching decisions will itself be undermined by the conclusion from which you are trying to escape.

This has grievous repercussions. Recall that one of the pillars of our Korsgaardian normative framework is that upon becoming aware of an impulse to do or believe something, you *cannot* but pose the question of what to do, which commits you to deliberating and making a decision about whether to do or believe that thing (see section VII above). For as long as you are self-conscious, then, you are compelled to deliberate and reach conclusions. But for as long as you believe that your epistemic competence is poor and hence that you are poor at deliberating, your deliberations and conclusions are perpetually undermined. Succumbing to stereotype threat—coming to accept its core belief—snares you in a loop of self-sabotage.

That analysis of the core belief also sheds light on the additional normative features of succumbing to stereotype threat, as well as on why they should cluster together. Beginning with your agential input, the constructivist framework shows that that input is intrinsic to the degenerate structure of the core

belief. Your agency consists in the exercise of your authority to yourself, and it is precisely because succumbing to stereotype threat involves your own authority to yourself that the content of the core belief is effective in warping that authority. However, because exercising your authority to yourself is the fundamental requirement to which you are subject (see section VII above), constructivism does not blame you for it.

As for the phenomenology of succumbing to stereotype threat, I contend that it just is the individual's experience of the pernicious structure of the core belief. That structure has you pitted against yourself at the most fundamental level—the level of your self-value as the author of the answer to the question of what to do. If you are being attacked by the very thing you value most, it is not implausible to think that your self-esteem would be in distress. And if your fundamental self-value is being undermined by itself, your inextricable commitment to expressing that value in what you do will be persistently frustrated, even as you cannot rescind that commitment. This will blight your ability to see yourself reflected in your decisions and, with it, your ability to see yourself as able to engage with the world in ways that are meaningful to you. Self-doubt too will take hold—if you are not good at providing the answer to the question of what to do, can you trust yourself about anything at all, including an answer to this question?

Finally, there is the sense in which you are a victim. We have seen that the wrong of succumbing to stereotype threat lies in the relation between the authority of the core belief, on the one hand, and the belief's content, on the other. We have also seen that of these two elements, the first is beyond reproach: manifesting your authority to yourself is the most fundamental requirement to which you are subject (see section VII above). This cannot be what you are a victim of. What about the content of the belief?

A key feature of the content of the core belief is that it fits so readily into your epistemic landscape. We saw that what makes your epistemic landscape so welcoming to the content of the core belief is the pervasiveness of the stereotypes in question. We also saw that stereotypes are pervasive in your epistemic landscape, both because they are relentlessly reinforced by your surroundings as you are growing up and beyond, and because the individuals that first convey and reiterate these messages to you are, at that time, authoritative to you. In short, it is because of the way your epistemic environment shapes your own epistemic landscape that propositions stating the truth of the stereotypes fit better within it than not. In succumbing to stereotype threat, then, you are a victim of your epistemic environment.

It looks like constructivism delivers the picture of succumbing of stereotype threat we seek—that is, a picture that illuminates its various normative

features and how those features hang together to form a single phenomenon. The wrong of succumbing to stereotype threat is the corrupt relation in which you stand to yourself, a relation contained in the self-subversive structure of the core belief. Your agential input is essential in that structure, but you carry no blame for it, as you abide by the fundamental requirement to which you are subject. The phenomenology of succumbing to stereotype threat is just your experience of that baleful relation to yourself. And in all of this, you are a victim of your epistemic environment, for that environment creates the crucial conditions on the basis of which you are normatively compelled to accept the core belief.

Before we settle on that conclusion, we must address a couple of reservations about the out-and-out acquittal I have passed on any blame of you. We might wonder whether you really could not avoid succumbing to stereotype threat; and we might have questions about your place within the epistemic environment of which you are a victim. The first misgiving can be put to bed easily enough. The second requires a more delicate treatment: it forces us to make certain concessions, and it signals where our research should go next. I take them in turn.

x

We might think that it is in your hands to avoid succumbing to stereotype threat. Could you not simply refuse to accept the core belief? Or steer clear of those considerations that forewarn of the fateful conclusion? The problem with those two options is that they each involve undermining your agency too, for they each constitute a violation of the requirement to provide a unified conception of the world. To violate that requirement is to violate your own epistemic authority over yourself and, with it, the fundamental requirement to which you are subject (see section vii above). It is, on this account, to act wrongly.

Furthermore, purposefully avoiding the conclusion that best unifies your conception of the world precludes the expression of your authority over yourself in its practical guise too. This is because, as we saw earlier (in section vii), your self-conception as the author of your answer to the practical deliberative question presupposes that you be able to govern yourself by your decisions, which itself presupposes that your conception of the world be unified—that it present to you a picture of things in the world and their relations in a way that enables you to act in it. If your conception of the world is willingly disjointed, you undercut your self-conception as being able to govern your actions by your decisions. And so when deliberating about what to do, you cannot reflect in your deliberation the primacy of the value of that self-conception. Since

reflecting the primacy of that self-conception is how you count as the author of your decisions, and being the author of your decisions is how you meet the fundamental requirement to which you are subject, it follows that in adopting this course of action, you would once again be acting wrongly.

Willingly shunning the core belief is not a credible alternative to succumbing to stereotype threat, for it involves violating your authority to yourself, both in its practical and in its epistemic iterations. As far as this objection goes, you remain blameless for succumbing to stereotype threat. Let us look at the next concern, which is to do with your place in your epistemic environment.

XI

The concern here might be put thus: While clearly you are not responsible for the epistemic environment that shaped your early epistemic development, in the time since you attained epistemic maturity, might you not have contributed to the upkeep of that epistemic environment, including the stereotype of Marchers? And if you have, are you blameworthy for it? And if you are, are you not therefore to blame for succumbing to stereotype threat?

That you contribute to the upkeep of your epistemic environment and, with it, to the perpetuation of the stereotype of Marchers is pretty much inevitable. A full and detailed picture of how this happens is beyond the scope of this article, but for our purposes, it suffices to underline that broadly speaking, you contribute to the perpetuation of the stereotype of Marchers simply by conforming to it; and if you are to be a functioning member of your society, you cannot but conform to it, at least to some extent. As Haslanger explains, societies are shaped to accommodate the purported natures of the various groups they tabulate.[50] Stereotypes function, among other things, as shortcuts for the prevailing understandings of the nature of each of those groups. A society's practices, structures, institutions, and general mores are tailored to fit the capabilities and limitations assigned to the chief stereotyped groups in that society. This all but forces individuals to operate within the spaces provided for them. If you transgress, you might be regarded as an abomination or simply not understood. Either way, your ability to go about your business is hampered. But in staying within the spaces assigned to you, you confirm the underlying view that those spaces are right for you, that they provide just what you need given what you are, and that the actions you perform within them reflect those facts.[51] For as

50 Haslanger, "Objectivity, Epistemic Objectification, and Oppression."

51 One of Haslanger's themes in her article "Objectivity, Epistemic Objectification, and Oppression" is the epistemic failings involved in inferring confirmation of the stereotype

long as you are a member of your community, then in all likelihood, you conform to and thus confirm the stereotype of Marchers, if only some of the time.

But this is not to say that you are to blame for it. Indeed, we might think that you are not even responsible for it—that is, that confirming the stereotype is not an action that can be attributed to you. On our constructivist framework, as on other accounts of action, something you do is not considered an *action* of yours unless a conception of it features in your decision to do what you do. For example, if last night you went to the kitchen to drink some water, and unbeknownst to you, your turning the lights on alerted a prowler to your being home, the alerting of the prowler is not considered an action of yours, while turning the lights on, grabbing the glass, turning the tap on, etc. are considered actions of yours.[52] We might think that your confirming the stereotype as you simply go about your business is akin to your alerting the burglar to your presence: you have no idea that your actions are having those effects.

However, if partaking in actions, activities, and practices that conform to the stereotype serve to confirm the stereotype to members of your community, since you are a member of your community, they must confirm the stereotype to you too. As such, you must have some awareness that that is one of the things you are doing. This is not a glib logical point. If you are a member of a community, you must have a sense that on the whole, you share a conception of the world with it. This sense of a shared understanding feeds into the feeling of kinship inherent in communities, and it also enables us to coordinate our actions. But that we share a conception of the world is something that wants being communicated over and over again, not least because, as we saw earlier, conceptions of the world are constantly in some degree of flux (see section VII above). In our interactions with each other, we tap (tacitly or overtly) into relevant meanings to request and to provide reassurances that we do see the world in broadly the same ways. Conformity to stereotypes is part of this tapestry of meanings: by confirming a shared view of what Marchers are like, you indicate that you share your community's conception of the world. In being a member of your community, you must be alive to these meanings, just like everyone else around you. At some level, then, if only subliminally, you must register the fact that one of the things you do as you interact in your community is conform to the stereotype often *in order to* confirm it.[53]

from conformity to it. I cannot put her insights to work here. Although those insights would enrich my final conclusion, they would not alter it fundamentally.

52 Davidson, "Actions, Reasons, and Causes," 4–5.

53 Another interesting aspect of this discussion that I must also forgo is how we negotiate and communicate differences between different individuals' conceptions of the world within the same community.

There are more flagrant cases too—cases in which the confirmation of stereotypes is front and center of what we do. Rituals and formal occasions have ostensive functions, e.g., to unite a couple in marriage, to confer an award. But just as obvious if not explicitly stated is their function to reaffirm a community's values. This often comes in the requirement for participants to perform exaggerated versions of some of the stereotypes that apply to them. It would be hard to insist that participants in these rituals are unaware that they are confirming the relevant stereotypes just as they dedicate inordinate amounts of time, effort, expense, and discomfort to making sure they do.

If those considerations are right, you are, at least to some degree, responsible for confirming the stereotype of Marchers. But this still does not tell us whether you are to blame for it.[54] For you to be blameworthy, you have to have acted wrongly. And you act wrongly only if your action does not reflect your values, specifically the primacy of your self-value as the author of your decisions. Does your engagement in activities that confirm the stereotype, such as those outlined above, reflect your values?

To an important degree, they might well do. Being a member of your community might be one of the things you most value about yourself and your life. If so, partaking in the activities that reinforce your community's ties, structures, and mores must be an expression of that self-conception. Similarly, you might value being a good neighbor, friend, or sister. Again, manifesting the commitments integral to those self-conceptions makes use of the communicative resources available—resources that, as we have seen, reflect the prevalent conception of the social world, including the dominant stereotypes in it.[55] Furthermore, those self-conceptions are themselves inflected by the dominant stereotypes. You are not just someone's neighbor, friend, or sister; you are a Marcher neighbor, a Marcher friend, a Marcher sister. Even if you have reservations about some of the values affirmed in the expression of any of those self-conceptions, you might go along with them anyway. You might relish the prospect of staging a protest against Marcher stereotypes at your friend's baby shower, but the thought of what that would do to your friend and to your friendship might make you recoil, so you may very well turn up to the shindig and behave as expected of you. On occasion, you might sacrifice the requirements of those cherished relationships for "the cause," but you cannot do it every time. If you did, not only might you find that requests for your company dry up, thus removing further opportunities to continue your revolt

54 Not, that is, unless we were consequentialists, but that would be anathema to the constructivist framework here advanced.

55 Midgley, *Beast and Man*, 310.

against the given stereotypes, but you would in effect relinquish those treasured self-conceptions—a member of your community, a friend, a sister—because you would cease giving expression to their constitutive commitments.

Confirming a stereotype, then, can quite possibly reflect self-conceptions that you hold dear. But as we saw earlier (see section VII above), that is not enough for your actions to be right. Of even greater value than any of those self-conceptions is your self-conception as your own authority—the self-conception that underpins all others. Therefore, its supremacy must be reflected in your decisions if your decisions are to count as right (see section VII above). If the stereotype of Marchers challenges that value, decisions to participate in activities that knowingly confirm that stereotype meet the conditions of wrong action, and as such, strictly speaking, you are blameworthy for them. After all, then, our constructivist approach does not absolve you of all blame.

We rejected the truth norm of belief and virtue epistemology partly because they blame the victim of succumbing to stereotype threat for succumbing to stereotype threat. Are we to think that constructivism no better? I think not. What blame is assigned to you here is mollified by some of the factors we have just highlighted. Reflecting your fundamental value as your own authority to yourself and thus refusing to confirm the stereotype would be tantamount to renouncing membership of your community and self-conceptions such as a friend, sister, or mother; it would involve parting with the people you love and who love you, and with the world you know. Certainly, the choice is yours: either remain a member of your community and keep the relationships that nourish your life at the cost of denying the full value of your authority to yourself, or stay true to yourself and go it alone. But we all, members of your community, face that choice, and we all opt for the same option, the wrong one: we stay.[56] Therefore, we are all to blame for maintaining our epistemic environment and the stereotypes in it—however they affect us. In that case, blaming *you* for it does not set you apart from anyone else; it does not single you out as *the* culprit of your succumbing to stereotype threat. The blame for

56 Emigrating to another community might not cut it either, at least not if that swapped one set of stereotypes for another. That is because there is a question as to whether one could ever willingly comply to a stereotype without violating one's authority to oneself. On the face of it, if one's actions are selected on the basis that they meet the stereotype's expectations rather than on the basis that they reflect one's values, it seems that one is not abiding by one's principal imperative; hence, one is once more selling oneself short. This needs more exploration than I can embark upon here, but I hope these considerations suffice to show that the options for social creatures like us of staying true to our ultimate authority are slim. My thanks to an anonymous referee for pressing me on this.

your succumbing to stereotype threat is shared with the rest of your community, past and present.[57]

This contrasts starkly with both the truth norm of belief and virtue epistemology, under both of which you are directly blamable for the particular instance of your succumbing to stereotype threat: the blame for succumbing to stereotype threat is yours alone, and had you not failed in the relevant ways (by accepting a false belief under the truth norm of belief or by being insufficiently virtuous under virtue epistemology), other things being equal, you would not have succumbed to stereotype threat. Any degree of blame for the circumstances of which you are a victim is uncomfortable, but I submit that the more nuanced blame assigned to you by constructivism is more tolerable than that apportioned by rival accounts.

XII

I conclude that a constructivist approach provides us with the unified understanding of succumbing to stereotype threat that we seek. Your epistemic environment, persistently and in countless ways, instills in you the message that in being a Marcher, your epistemic competence is poor. On the strength of sustained reinforcement, that message becomes a key plank in your epistemic economy, thus playing a decisive role in the acquisition of new beliefs. When facing the proposition that in being a Marcher, your epistemic competence is poor, you face a proposition that is quite at home within your belief system. Accepting that proposition on the basis that doing so makes for a more unified conception of the world than not is the right thing to do. But that proposition corrupts the very dynamic of belief formation: it negates the authority that underpins your accepting that proposition in the first place. Since that authority is fundamental to you as an agent, its use to undermine itself is tantamount to turning yourself against yourself. You experience that insidious contortion as loss of meaning and of self-esteem, with low expectations and low motivation. You are fully agentially engaged in this episode—that is whence its force comes. The dreaded proposition would not have the reach it has if it were presented by a source for whom you had little regard. Yet you are not to blame for it directly—that is, you are not to blame for the particular event of your succumbing to stereotype threat. Direct blame lies with your epistemic environment, and that environment is what you are a victim of. However, you are part of that environment too, and that means that you are an active participant

57 Of course, the distribution of blame is surely not even, and this is something that requires attention. But our interest here is simply whether you can be blameless for succumbing to stereotype threat, not how much blame you shoulder.

in it. This all but necessarily involves confirming its central ontological commitments, including those codified in the stereotype of Marchers. You do this in a way that accrues blame, but a number of factors mitigate the extent to which that blame translates into blame for succumbing to stereotype threat: the connection between the offending activities and the particular instance of your succumbing to stereotype threat is diffuse; your participation in those activities puts you on a level with the rest of your community; and the cost of refusing participation would be self-imposed ostracism.

However, that conclusion exposes its own limitations. I set out above to examine the internal normative structure of stereotype threat in itself by focusing on the phenomenon of succumbing to stereotype threat. But our study has revealed that the normative structure of succumbing to stereotype threat is not confined to a discrete phenomenon. Rather, one's social environment, critically and unavoidably, seeps into it. You are to blame because you are part of the community that shaped and maintained your epistemic landscape such that it plays a pivotal role in rendering your succumbing to stereotype threat both compelling and correct. I said in section XI that given the scope of this article, I could but dip my toes into that social dimension. But the ensuing discussion has made clear, if it was not already, that further probing into that social grounding—into the extent to which you are complicit or conscript, into the extent to which that grounding makes you or debases you—is indispensable for a full understanding of succumbing to stereotype threat and thus a full understanding of stereotype threat.

I propose that there is a readily available entry point into that further research: the notion of *alienation* as understood by Marx. Marx's notion of alienation is embedded in a rich tradition that pays close attention to the porous relationship between the individual and her environment—a complicated and dynamic relation, both causal and constitutive—without losing sight of the relation the individual has to herself. The notion of alienation is apposite because our picture of succumbing to stereotype threat neatly falls under that notion. To make my case, I look at the broad structure of alienation. This can be only brief at this point, but I hope it suffices to show its promise.

Marx's notion of alienation underwent a range of mutations, but its leitmotif is a kind of "hostile relationship" between things that belong together, including, crucially, the relation in which the individual stands to herself.[58] This hostile relationship in which the individual stands to herself is a corruption of the way in which the individual properly engages with the world.

58 Wood, *Karl Marx*, 3, 4, 6; Wolff, *Why Read Marx Today?* 29; and Marx, "Economic and Philosophic Manuscripts," 64.

The idea here is that the individual's fundamental values are comprised in her self-conception as a human being.[59] As she interacts with her environment, the individual has a *need* to express those values, a need for *self-actualization*.[60] Proper engagement with the world is the satisfaction of that need. Alienation is its frustration—it is the condition of going about one's life in a way in which one "does not affirm himself . . . but denies himself," in which "the realization of his nature [appears] as the diminution of his life."[61] In this way, "alienated individuals are in some sense separated from, at odds with, or hostile to themselves," their activity turned against them as a kind of "torment."[62] In an alienated condition, the individual suffers lack of self-worth and "experiences life as empty, meaningless and absurd."[63] The background setting for alienation, Marx thought, is "a complex interconnection between . . . various ills and irrationalities" present in society.[64] The alienated individual is participant in maintaining the circumstances of which she is victim, yet altering those circumstances is beyond her power.[65]

The parallels between this broad characterization of alienation and our constructivist diagnosis of succumbing to stereotype threat are plain. The hostile relationship to yourself that is the hallmark of alienation appears in succumbing to stereotype threat in your acceptance of the core belief, in employing your authority to undermine itself. As in alienation, that relationship arises in a corruption of your proper epistemic engagement with the world. Your proper epistemic engagement with the world is one in which you exercise the supreme value of yourself as your own authority. And exercising that value is a need you have in virtue of being the kind of creature you are—it is that need, recall, that is the source of normativity (see section VII above). Again in consonance with alienation, that need is frustrated when you undermine your authority to yourself. The experience of that relationship as lack of self-esteem and associated loss of meaning and of motivation mirrors the experience of alienation as one of meaninglessness, emptiness, and self-worthlessness. The society's ills and irrationalities that foster alienation are found, in relation to succumbing to stereotype threat, in the form of distorted conceptualizations of individuals, together with the social structures that demand compliance with

59 Wood, *Karl Marx*, 19–20.

60 Wood, *Karl Marx*, 22, 26–30.

61 Wood, *Karl Marx*, 23; and Marx, "Economic and Philosophic Manuscripts," 61–62, and "Excerpt Notes of 1844," 46.

62 Wood, *Karl Marx*, 8; and Marx, "Excerpt Notes of 1844," 46.

63 Wood, *Karl Marx*, 21, 8.

64 Wood, *Karl Marx*, 7.

65 Wood, *Karl Marx*, 8, 14; and Marx, "Economic and Philosophic Manuscripts," 66.

them. Also replicating the setting in alienation, so long as you remain in your society, you cannot but contribute to those maladies. Yet you have no power to change them anywhere near the degree where they would no longer facilitate instances of succumbing to stereotype threat. The phenomenon of succumbing to stereotype threat, I conclude, is a type of alienation.

The notion of alienation names the wrong of succumbing to stereotype threat and thus the substantive threat of stereotype threat.[66] It is already part of our analysis of succumbing to stereotype threat and thus of stereotype threat, and so it is well poised to help us delve deeper into the social dimension of those phenomena.

XIII

My aim in this article has been to expand our understanding of the normative charge of stereotype threat. To do so, I have foregrounded the phenomenon of succumbing to stereotype threat in cases of stereotype threat that present us with the threat of epistemic self-objectification—that is, cases in which (1) the stereotype asserts that in virtue of belonging to the stereotyped group in question, your epistemic competence is poor and (2) the threat pertains to confirming the stereotype at least to yourself.[67] I have shown that succumbing to stereotype threat is the formal threat of stereotype threat. A constructivist approach has exposed the fundamental wrong of succumbing to stereotype threat to be the individual's own subversion of her own authority to herself, and it has elucidated how all other normative strands of succumbing to stereotype threat are attached to that wrong. This has led us to the individual's social environment, where we have gauged a rough overview of some of the involuted and pervasive ways in which the individual and her environment feed into each other. I have argued that this analysis of succumbing to stereotype threat corresponds to Marx's notion of alienation. We are thus afforded the conclusion that the substantive threat of stereotype threat in the cases at hand is alienation.

That analysis has also belied my initial aim of confining my research to the inner normative structure of stereotype threat. We have seen that the phenomenon of succumbing to stereotype threat (and thus of stereotype threat) is not a discrete unit but is rather inexorably tied to a social environment. This discovery forces me to end my study on a pending task: a fuller understanding

66 The advantage of being able to name the wrong of succumbing to stereotype threat and the threat of stereotype threat is hermeneutical. In naming what is a variegated normative phenomenon, we can more easily recognize it, talk about it, and address it. See Fricker, *Epistemic Injustice*.

67 Haslanger, "Objectivity, Epistemic Objectification, and Oppression," 280.

of the phenomenon of succumbing to stereotype threat (and thus of stereotype threat) requires a fuller enquiry into the ontological and normative fluctuations of that social terrain. I have briefly argued that the notion of alienation is well placed to help us with that extra task.[68]

Durham University
anabarandalla@me.com

REFERENCES

Alfano, Mark. "Stereotype Threat and Intellectual Virtue." In *Naturalizing Virtue*, edited by Owen Flanagan and Abrol Fairweather. Cambridge University Press, 2014.

Antony, Louise. "Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy?" *Journal of Social Philosophy* 43, no. 3 (2012): 227–55.

Bagnoli, Carla. "Moral Constructivism: A Phenomenological Argument." *Topoi* 21, nos. 1–2 (2002): 125–38.

———. "Self-Deception and Agential Authority: Constitutivist Account." *Humana Mente* 5, no. 20 (2012): 93–116.

Battaly, Heather. "Virtue Epistemology." *Philosophy Compass* 3, no. 4 (2008): 639–63.

Beebee, Helen. "Women and Deviance in Philosophy." In *Women in Philosophy: What Needs to Change?* edited by Katrina Hutchison and Fiona Jenkins. Oxford University Press, 2013.

Blum, Lawrence. "The Too Minimal Political, Moral, and Civic Dimension of Claude Steele's 'Stereotype Threat' Paradigm." In *Implicit Bias and Philosophy*, vol. 2, edited by Michael Brownstein and Jennifer Saul. Oxford University Press, 2016.

Bortolan, Anna. "Self-Esteem and Ethics: A Phenomenological View." *Hypatia* 33, no. 1 (2018): 56–72.

Davidson, Donald. "Actions, Reasons, and Causes." In *Essays on Actions and Events*, 2nd ed. Oxford University Press, 1963.

Fricker, Miranda. *Epistemic Injustice*. Oxford University Press, 2007.

Goguen, Stacy. "Expanding Our Picture of Stereotype Threat." *Minds Online* (blog), September 10, 2016. http://mindsonline.philosophyofbrains.com/2016/2016-2/expanding-our-picture-of-stereotype-threat/.

———. "Stereotype Threat, Epistemic Injustice, and Rationality." In *Implicit*

68 My thanks to two anonymous referees for their comments and encouragement.

Bias and Philosophy, vol. 1, edited by Michael Brownstein and Jennifer Saul. Oxford University Press, 2016.

Haslanger, Sally. "Objectivity, Epistemic Objectification, and Oppression." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus. Routledge, 2017.

Korsgaard, Christine M. "The Activity of Reason." *Proceedings and Addresses of the American Philosophical Association* 83, no. 2 (2009): 23–43.

———. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press, 2009.

———. *The Sources of Normativity*. Cambridge University Press, 1996.

Mallon, Ron. "Stereotype Threat and Persons." In *Implicit Bias and Philosophy*, vol. 1, edited by Michael Brownstein and Jennifer Saul. Oxford University Press, 2016.

Marx, Karl. "Economic and Philosophic Manuscripts." In *Karl Marx: Selected Writings*, edited by Lawrence H. Simon. Hackett Publishing, 1994.

———. "Excerpt Notes of 1844." In *Karl Marx: Selected Writings*, edited by Lawrence H. Simon. Hackett Publishing, 1994.

McHugh, Conor. "The Truth Norm of Belief." *Pacific Philosophical Quarterly* 93, no. 1 (2012): 8–30.

McKinnon, Rachel. "Stereotype Threat and Attributional Ambiguity for Trans Women." *Hypatia* 29, no. 4 (2014): 858–72.

Midgley, Mary. *Beast and Man: The Roots of Human Nature*. 1st ed. Routledge, 2002.

Nguyen, Hannah Hanh D., and Ann Marie Ryan. "Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence." *Journal of Applied Psychology* 93, no. 6 (2008): 1314–34.

Ratcliffe, Matthew James. "Existential Feeling and Narrative." In *Funktionen Des Lebendigen*, edited by T. Breyer and O. Mueller. De Gruyter, 2016.

Rawls, John. *A Theory of Justice*. Rev. ed. Harvard University Press, 2005.

Saul, Jennifer. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy: What Needs To Change?*, edited by Katrina Hutchison and Fiona Jenkins. Oxford University Press, 2013.

Schouten, Gina. "The Stereotype Threat Hypothesis: An Assessment from the Philosopher's Armchair, for the Philosopher's Classroom." *Hypatia* 30, no. 2 (2015): 450–66.

Shapiro, Jenessa, Joshua Aronson, and Matthew S. McGlone. "Stereotype Threat." In *Handbook of Prejudice, Stereotyping, and Discrimination*, 2nd ed., edited by Todd D. Nelson. Psychology Press, 2016.

Steele, Claude M. *Whistling Vivaldi: How Stereotypes Affect Us and What We*

Can Do. W. W. Norton, 2011.
Street, Sharon. "Coming to Terms with Contingency: Humean Constructivism About Practical Reason." In *Constructivism in Practical Philosophy*, edited by James Lenman and Yonatan Shemmer. Oxford University Press, 2012.
Taylor, Charles. "Rationality." In *Rationality and Relativism*, edited by Martin Hollis and Steven Lukes. MIT Press, 1982.
Valian, Virginia. *Why So Slow? The Advancement of Women*. MIT Press, 1998.
Velleman, J. David. *How We Get Along*. Cambridge University Press, 2009.
Wolff, Jonathan. *Why Read Marx Today?* Oxford University Press, 2003.
Wood, Allen W. *Karl Marx*. Routledge, 1984.
Zagzebski, Linda Trinkaus. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press, 1996.
Zangwill, Nick. "The Normativity of the Mental." *Philosophical Explorations* 8, no. 1 (2005): 1–19.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.4135

WHY (AND HOW) STATES PROTECT CHILDREN

Nicholas Hadsell

When should the state intervene in the parent-child relationship? Surely it should do so in cases of severe abuse and neglect. However, many theorists have argued that children's rights go beyond the mere avoidance of abuse and neglect to include the right to a carefree childhood, quality education, and even love.[1] Whatever the full content of children's rights is, the state intervenes when only a small subset of those rights is violated. Is this justified?

Moreover, the strong set of rights that parents enjoy in the status quo prevents the state from adequately enforcing children's rights against abuse and neglect. In particular, the bar to becoming a legal parent in the United States is surprisingly low.[2] Regardless of the fitness and living conditions of biological parents, the state frequently places newborns into legal parent-child relationships with them. The state disregards the fact that the same biological parents may have already lost custody of their other children or may be suffering from severe drug addiction or mental illness at the time of birth. Instead of doing anything about these factors, the state compels newborns to be in a legal family relationship with the biological parents simply because they are the biological parents. While children are somewhat resilient, findings from adolescent developmental psychology show that the effects of parental abuse and neglect are significant and long-lasting.[3]

The state currently has a wait-and-see approach to intervention; once it has evidence of sufficiently extreme abuse and neglect, it has grounds to act. However, this evidence is notoriously hard to procure. While child protective services reports two million substantiated cases of abuse and neglect every year,

1 For a carefree childhood, see Ferracioli, *Parenting and the Goods of Childhood*, part 2. For education, see Brighouse and Swift, *Family Values*; and Clayton, *Justice and Legitimacy in Upbringing*. For love, see Liao, *The Right to be Loved*; and Ferracioli, "The State's Duty to Ensure Children Are Loved."

2 In support of what follows in this paragraph, see Dwyer, *Liberal Child Welfare Policy and Its Destruction of Black Lives*, 66; and Dwyer, "The Child Protection Pretense," part 1.

3 Dwyer, "The Child Protection Pretense," part 2.

the actual number is likely three times greater.[4] With just a little imagination, it is not hard to see that there is likely much more parental abuse and neglect than the state finds: parents are usually fully developed adults, while children are incredibly vulnerable to their parents for most of their childhood. Such abuse and neglect are especially likely to go undetected in contexts where children have no regular contact with anyone besides their parents.[5] The state is supposed to protect children from abuse and neglect, but it is incredibly difficult for it to do so. Is this justified?

To make any headway on these questions, we first need a theory of the state. As David Archard has argued, "what is regarded as the proper role for the state in the protection of children's interests will be crucially influenced by how the state itself is viewed."[6] That is, the answer to these questions about whether status quo family law is justified depends on what we think the state is supposed to do in our communities.[7] For example, if philosophical anarchism is true—i.e., the state lacks the right to be obeyed—then the state has little role in familial affairs beyond the general duties of rescue to make sure children do not suffer severe abuse and neglect. Conversely, if some variant of Rawlsian liberalism is true—i.e., the state is responsible for making sure the basic structure of society is just, *and* the family is part of that basic structure—then the state might have more leeway in regulating familial affairs.[8]

One goal of this article is to advance a broadly "Kantian-republican" theory of the state, one in which the state is a necessary precondition to any minimally just society.[9] If this theory is correct, there are significant theoretical upshots for what we should say about the family and the laws that should regulate it. In particular, this theory undermines the long-standing view of the family as

4 LaFollette, "Parental Licensing Revisited," 333.

5 For a good discussion about the vulnerability of children, see Archard, "Child Abuse." Regarding why this spells trouble for monopolies of care, see also Gheaus, "Children's Vulnerability and Legitimate Authority over Children."

6 Archard, *Children*, 153.

7 Luara Ferracioli makes a similar point when she observes that liberals often assume states are required "to act paternalistically toward children" without explaining the requirement ("Citizenship for Children," 2868).

8 For example, see Neufield, "Coercion, the Basic Structure, and the Family."

9 I get this term from Kolodny, "Being Under the Power of Others." While I am mainly focused on the Kantian part of the term, I use it here because I think Kant counts as a kind of republican given his central concern with domination. (See Ripstein, *Force and Freedom*, 42–43.) For examples of the Kantians I am drawing from, see Ripstein, *Force and Freedom*; Ebels-Duggan, "Moral Community"; and Varden, "Kant's Non-Voluntarist Conception of Political Obligations"; and Pallikkathayil, "Deriving Morality from Politics."

a *prepolitical* institution that can exist rightfully in the state of nature.[10] The practical outcome of the prepolitical view is generally that the state has a minimal role in familial affairs, intervening only in extreme cases.[11] In contrast, the view I defend sets a lower bar for state intervention. Additionally, the Kantian-republican view undermines any account of the family that says it is merely a convention of the state, rendering the family nonexistent before any positive state action.[12] If this conventional view of the family is true, then the state has a larger role in creating and sustaining families under its jurisdiction. Comparatively, my view of the state's role is not as expansive.

Between these views, I argue for a different theory of the state-family relationship: families can exist in the state of nature (*contra* the conventionalist position), but they cannot do so nondefectively (*contra* the prepolitical position). In particular, I build on recent developments of Kantian-republicanism, especially Helga Varden's early work, to argue that children in the state of nature necessarily suffer domination.[13] The reason is that children have a set of rights against their parents to be treated in certain ways, but nobody has the authority to enforce most of those rights on behalf of the child against their parents. Here, parents are akin to enslavers who have full rights over persons and may use their powers benevolently but are nonetheless dominating children who have no way to enforce their rights against the parents.[14] However, if there were a public state entrusted with the power to legislate, enforce, and adjudicate family law, then children would not necessarily be subject to slave-like relationships with their parents. Because everyone has a duty to respect everyone's rights, and the state is the only context in which we can do so for children, everyone must submit to a public state that enforces children's rights.

If the foregoing is correct, there are also significant practical implications for what the state can permissibly do to regulate the family. Philosophers, family lawyers, and social workers have extensively debated proposals such as parental

10 For example, see Moschella, *To Whom Do Children Belong?*; and Locke, *Two Treatises of Government*.

11 For example, see Moschella, *To Whom Do Children Belong?* 67–69.

12 For a discussion of this view, see Shelby, *Dark Ghettos*, 144–45.

13 I first came across the possibility that children are necessarily slaves under their parents in Helga Varden's analysis of status relationships in Varden, "Kant's Non-Voluntarist Conception of Political Obligations," 21–24. Here, I aim to flesh out this idea, connect it to the ethics of parenthood literature, and examine how the connection bears on child welfare policies like licensing and monitoring.

14 I follow Arthur Ripstein in using 'domination' interchangeably with 'wrongdoing', which is the term we use when one person violates another's right to independence. See Ripstein, *Force and Freedom*, ch. 2, sec. 3, where he writes that "wrongdoing takes the form of domination" (42).

licensing and parental monitoring to safeguard children's rights against parental misconduct. Proponents of licensing argue that the state should use evaluative tools to determine whether parents are competent enough to care for their children.[15] Whereas parental licensing regulates the initiation of parent-child relationships, monitoring is supposed to regulate already existing parent-child relationships through occasional contact between children and state representatives.[16] If my view is correct, then there are probably no good *in-principle* objections to either of these policies. If we have reason to forego either of them, it is because they do not adequately secure children's rights without domination in practice for a variety of contingent, empirical reasons. This means we have a conditional defense of both policies: *if* they can permissibly safeguard children's rights without domination, states may implement them. In-principle objections such as "it is not within the state's purview to regulate the family in these ways" do not work.[17]

Here is the plan for the rest of the article. In section 1, I begin my account of the state's relationship to the family by showing how families in the state of nature necessarily dominate children. In section 2, I discuss the solution to this predicament: the community establishes a public state capable of and responsible for addressing the defects of the state of nature. In section 3, I show how this solution might look in practice by considering the viability of schemes like licensing and monitoring. I conclude that whether a state should enforce either of these policies depends on prudential considerations about whether these policies have a good chance at protecting children's rights without domination.

Before beginning, I should answer an important question about the argumentative strategy I use. Why do we need to consider what happens to children in the state of nature, as we do in the next section, rather than simply looking at a given society and saying what the state must do to protect children from domination? The reason is that while the latter strategy might reveal that the state can play a privileged role under certain conditions, the former strategy can show something stronger: it explains why it is uniquely the state's job to prevent domination. It is one thing to say domination is something we should prevent and that, in one context, the state happens to be best suited to prevent it; it is another to say it is uniquely the state's job to prevent it in any conditions

15 Proponents of licensing include LaFollette, "Parental Licensing Revisited"; and Kianpour, "The Kids Aren't Alright."

16 Proponents of some sort of monitoring include Archard, "Child Abuse"; and De Wispelaere and Weinstock, "Licensing Parents to Protect Our Children?"

17 This view is therefore compatible with practical objections to licensing that say it is too costly to enforce. For examples of such objections, see Archard, "Child Abuse," 190–91; and Freiman, "Against Parental Licensing."

where that domination could occur. For many social contract theorists, the state of nature helps identify broad principles that delineate the limits of political authority and, consequently, how expansive the state's role should be in preventing domination.[18] As we see below, different descriptions of the state of nature yield different results concerning the size of this role: views that are more optimistic about the determinacy of natural rights generally afford the state a more restrictive role than views that are much more pessimistic about the determinacy of natural rights. As I show in section 2, this divergence matters for which sorts of child welfare policies the state has the authority to enact.

1. THE STATE-FAMILY RELATIONSHIP

To understand how the state should relate to the family, we can start by thinking about the network of rights and obligations that generally characterize the parent-child relationship in a healthy case. Once we understand this network, we can see why familial relationships would be normatively defective in the state of nature—i.e., because children are necessarily subject to a slave-like relationship before their parents—and why the state is the only remedy.

1.1. The Normative Network of the Parent-Child Relationship

When a couple procreates, they bring a child into the world who innately has a set of rights against them. Kant, for example, claims that procreative parents "incur an obligation to make the child content with [the child's] condition so far as they can."[19] Parents incur this obligation because they have decided to bring a child into existence without the child's consent, which means they must make that decision a sufficiently good choice for the child.[20]

For Kant, what explains this requirement on parents is that the child, like any other human being, has an innate right to external freedom to set and pursue the ends they set for themselves.[21] However, whereas adults have

18 Most Kantian accounts of the state use the same methodology. For example, see Ripstein, *Force and Freedom*, ch. 6; and Stilz, *Liberal Loyalty*, ch. 2.

19 Kant, *Metaphysics of Morals*, 6:281.

20 Porter, "Why and How to Prefer a Causal Account of Parenthood," 194. Here, I assume the causal theory of obligation. While many (e.g., Porter) read Kant as endorsing this view, others (e.g., Ripstein, *Force and Freedom*, 72) do not. My goal here is to defend a theory about the enforcement of obligations, not one of how those obligations come about. Others who reject the theory may be able to substitute their preferred account.

21 "Children, as persons, have by their procreation an original innate (not acquired) right to the care of their parents until they are able to look after themselves, and they have this right directly by law (*lege*), that is, without any special act being required to establish this right" (Kant, *Metaphysics of Morals*, 6:280).

developed the requisite capacities to exercise this right to external freedom on their own, children come into the world without those capacities; they are instead totally dependent on competent adults to raise them in certain ways so that they eventually develop the capacities they need to live independent lives. When a child comes into the world, then, their innate right requires their procreators to make sure someone plays the parental role and raises them to become an independent person who can exercise their innate right—that is the only way the biological parents' procreative decision would prove sufficiently good for the child.

Importantly, a child's innate right does not necessarily obligate or entitle the procreators to raise the child themselves on its own.[22] The innate right requires procreators to ensure that *someone* raises the child toward independence, whether that is the procreators themselves or someone else. We can understand this difference by borrowing a helpful distinction from Lindsey Porter. When procreators bring a child into the world, they incur what Porter calls a *maker obligation*, which is an "obligation incurred in virtue of having caused the child to exist … and is, roughly, the obligation to make the child's existence a good one to the extent that one can."[23]

Fulfilling the maker obligation does not necessarily require a procreator to raise their child. Procreators can fulfill their maker obligations by ensuring that someone else, whom they have good reason to believe will give their child a good enough life, raises their offspring. There are in fact some cases in which a procreator *should* give up their child to someone else because the child would be significantly worse-off if the procreator raised them. Think, for example, of a procreator who knows they are too mentally unstable or addicted to life-ruining substances to meet the child's needs in the parental role adequately. However, if no one else can adequately care for the child, then the maker's obligation requires the procreator to step into the parental role and to care for the child to the best of their ability.

This leads to the second part of Porter's distinction. Once someone steps into the parental role over a child, they then incur a *parental obligation*, which is "the obligation incurred in virtue of taking on the social role of parent … and includes all those everyday care obligations we would normally attribute

22 I get this insight from Archard, "The Obligations and Responsibilities of Parenthood," 114. Strictly speaking, this is not Kant's original view. He argued that procreators' duties beget all the rights one needs to raise the child to maturity (*Metaphysics of Morals*, 6:281). But loosely speaking, this is probably still consistent with the spirit of his view, especially once we depart from standard cases to ones in which the procreators in question clearly should not raise their offspring.

23 Porter, "Why and How to Prefer a Causal Account of Parenthood," 196.

to the social parent."[24] Parental obligations attach to whoever ends up raising a child; they explain why one adult, and not another, wrongs a child if they fail to feed and clothe the child regularly. Distinctly parental obligations are entailed by the child's right to adequate care, enabling the child to grow into an independent person.

For a parent to meet these obligations, they must also possess the necessary rights to do so. Locke, for example, described parental authority as "a sort of rule and jurisdiction" or "temporary government" that parents possess over their children so that they can raise them into independence.[25] The rights that constitute such parental authority are what A. John Simmons calls "mandatory claim rights," which are "rights to do what we have a prior duty to do."[26] If children have a right against their parents to be raised toward independence, parents must possess the requisite authority over their children to fulfill those rights. Parents must, for example, be able to discipline their children appropriately if such discipline is required for children's proper development.

Kant similarly endorsed the principle, claiming that "*from this duty* there must *necessarily also arise the right* of parents to *manage* and develop the child, as long as he has not yet mastered the use of his members or of his understanding."[27] That is, parental obligations create parental rights to the child. However, Kant thought parental rights were circumscribed in an important way by children's innate rights. The reason is that a parental right is a species of what Kant calls a *status right*, which is the right one person gets "in a relationship in which one party is not in a position to consent either to the existence of that relationship or to modification of its terms."[28]

Standard cases of status rights include the rights that fiduciaries have to act on behalf of their conservators, the rights that teachers have to manage their students' schedules while they are in class, and, paradigmatically, the rights that parents have to manage and develop their children. In all cases, a status right holder has the right to make a variety of decisions for the person over whose property and behavior they have status right: a fiduciary may use their client's money to pay for the client's electric bills; a teacher may have their students sit outside to do a math lesson; and a parent may take their child to the doctor to treat a sickness. In this sense, having a status right vis-à-vis someone is having

24 Porter, "Why and How to Prefer a Causal Account of Parenthood," 193.

25 Simmons, *The Lockean Theory of Rights*, 178. See also Locke, *Two Treatises of Government*, 67–68.

26 Simmons, *The Lockean Theory of Rights*, 182.

27 Kant, *Metaphysics of Morals*, 6:281 (emphasis added).

28 Ripstein, *Force and Freedom*, 72.

a right to a person *akin to a thing*: it is like having a right to a thing insofar as the possessor can use the person for certain purposes and demand the person back if someone else takes them.[29] However, a status right is unlike a right to a thing insofar as the purposes that the status right holder has for their subject—whether that be a fiduciary client, a student, or a child—must respect that subject's innate right.[30] This is why, for example, a conservator cannot use their client's money to buy themself expensive jewelry; a teacher cannot force their students to mow the teacher's lawn; and a parent cannot sell their child for profit. Without this restriction, the rights that parents possess vis-à-vis their children—ones they acquired without the children's consent—would give parents the space to own children essentially like they own other pieces of property, without any rights against them. But children are different from things; they have innate rights, whereas property does not. As a result, the domain of things that we can do with our status rights is more restricted than the domain of things we can do with our property rights to inanimate, external objects.

We have, then, a relationship of parental authority: a child has an innate right that requires whoever parents that child to assist the child in their development toward independence. This requirement explains the parent's obligations and gives the parent a set of rights over the child that includes only those rights needed to meet their parental obligations.

1.2. The Slave Problem for Stateless Families

Within the state of nature, this relationship would *necessarily* wrong children, even if parents are perfectly virtuous.[31] On the one hand, the prepolitical normative network from the previous section shows that children have a set of weighty rights against their parents to be cared for in certain ways. On the other hand, within the state of nature, *these rights are impossible to rightfully enforce*.[32]

29 Kant, *Metaphysics of Morals*, 6:358.

30 Kant, *Metaphysics of Morals*, 6:277. See also Varden, "Kant's Non-Voluntarist Conception of Political Obligations," 265–66; and Ripstein, *Force and Freedom*, 79. Once a parent acquires a status right over a child, the child's rights obligate them. Here, the parent-child relationship is structurally the same as other status relationships. For example, fiduciaries are responsible for managing their clients' estates. Whether third-party citizens have duties of aid to these relationships is a separate question, though some Kantians deny these. For example, see Allais, "What Properly Belongs to Me"; and Hadsell, "Kant on Punishment and Poverty."

31 See Varden, "Kant's Non-Voluntarist Conception of Political Obligations," 21–24.

32 Domination is not about the enforcement of *moral* rights (e.g., my right against my friend to keep her promise), which are not enforceable and do not result in domination when violated. Instead, domination concerns only *juridical* rights, or "normative claim[s] against others that it is permissible to enforce, or to coerce others to respect" (Ebels-Duggan,

Therefore, children within the state of nature are necessarily subject to dominating relationships because most of their rights are unenforceable, no matter what happens empirically. We can see this in the following case.

> *Emily's Education*: Alex is raising his nine-year-old daughter, Emily, in a state of nature. Given the normative network of the parent-child relationship, Emily has a set of rights against Alex that requires him to ensure she is raised to become a mature, independent adult. Among these rights is the right to a minimally good education that equips her with a set of intellectual virtues that she will need to discern how to live a good life. However, Alex decides homeschooling is too much work given all the hunting and farming he has to do, and he is too skeptical of the local co-ops in the state of nature to place Emily in them. So Alex neither homeschools nor enrolls Emily in a co-op. Emily tells John, an unrelated third party who happens to be a teacher, about Alex's decision. John thinks he should intervene by enrolling Emily in his local co-op to ensure she receives a minimally good education.[33]

What can any of the parties rightfully do in this case? Not much; the case is incoherent as a matter of right because it inherits the following defects intrinsic to the Kantian state of nature.[34]

1. *Unilateralism*: When Alex claims a status right over Emily, he places everyone else, through his unilateral choice, under an enforceable obligation not to interfere with Emily's education without his consent.[35] But in the state of nature, Alex lacks the authority to create an enforceable entitlement like this

---

"Kant's Political Philosophy," 897). Of course, this invites an obvious question: If prepolitical juridical rights are unenforceable, how can children be dominated? In reply, I say that these prepolitical rights are juridical in that they are supposed to be enforceable, but they are defective because there is no prepolitical mechanism to rightfully enforce them. As Arthur Ripstein puts it, "They are all titles to coerce that nobody is entitled to enforce coercively" (*Force and Freedom*, 165). To make sense of this, we can compare the defective nature of prepolitical juridical rights to how Aristotelians analyze defects in creatures. Spiders with seven legs are still spiders, but some Aristotelians say that they are defective because it is in their nature to have eight legs. Similarly, juridical rights in the state of nature are defective because they are supposed to be enforceable, but they exist in an environment where there is no rightful enforcement mechanism. As a result, they are merely *provisional*. For more on provisional rights, see Ebels-Duggan, "Moral Community," 4–7; and Stone and Hasan, "What Is Provisional Right?" See also section 1.3 of this article.

33 I thank Alexander Pruss for this example.

34 The following is necessarily sketchy and relies heavily on Ebels-Duggan, "Kant's Political Philosophy," pt. 1.

35 Kant, *Metaphysics of Morals*, 6:255; and Ripstein, *Force and Freedom*, 153–54.

because doing so would violate everyone else's innate right.[36] Alex may claim he wants Emily to have a certain sort of education (or lack thereof), but he cannot unilaterally generate an enforceable obligation on others to avoid interference. This means that if John tries to intervene, Alex may resist, but John may also resist Alex's resistance. Neither party has a claim over the other.

2. *Indeterminacy*: Even though Emily formally has a right to education, there is no determinate answer within the state of nature that says how this right should be fulfilled materially.[37] The general right—i.e., a right to an education—is insufficient to apply itself to particular situations—i.e., Emily's right in her context. After all, there are various ways the right can go materially (e.g., John's co-op, an education oriented toward developing agricultural or industrial skills, a STEM-based education, etc.). To get from the general right to its particular application, there must be an authoritative determination over a certain set of these possible material manifestations of the right to an education. But given the problem of unilateralism, nobody within the state of nature has the authority to make such a determination. So when John and Alex disagree over the determination of Emily's right, neither has more fundamental standing to make the determination.[38] So the right is not enforceable because, trivially, it is indeterminate.

3. *Assurance*: Suppose, contrary to fact, that there is a determinate set of laws governing our rights in the state of nature that we can appeal to without unilaterally imposing our wills on others. Still, if I lack the assurance that my claims to things will be systematically enforced against others in the state of nature, then I cannot be obligated to respect others' similar claims.[39] This is not because respecting obligations without assurance is imprudent or detrimental to us in some way. Rather, Kant thinks everyone has rightful honor, which requires all to refrain from making themselves a means for others.[40] When

36 Kant, *Metaphysics of Morals*, 6:256.

37 Ripstein, *Force and Freedom*, 170.

38 Even if John and Alex end up agreeing on a material fulfillment of the right, the problem is still not solved. The reason is that neither of them individually has the authority to make a unilateral determination over Emily's education that binds others to not interfere with it. No matter if the unilateral choice maker is just John or John plus Alex, such unilateral choices are objectionable from the perspective of right. This is especially the case if Alex and John agree at one point, but then one of them changes their minds later. Emily's right to an education is contingent on the private choice of either party, which is bad because our rights are supposed to be based on universal, systematic laws, not on private choices. So only when the determination is *omnilateral*—i.e., it comes from an authoritative state that represents *all* its members—is when it becomes compatible with right. See Ripstein, *Force and Freedom*, 157.

39 Kant, *Metaphysics of Morals*, 6:255–56; Ebels-Duggan, "Moral Community," 5; and Pallikkathayil, "Persons and Bodies," 38.

40 Ripstein, *Force and Freedom*, 161; and Loriaux, "Kant on Social Justice."

we respect others' claims to things without assurance that others will respect our claims—an assurance we lack since there is no universal law in place that requires them to do so—we make ourselves a means.

The result of all these defects is that our case is, from a matter of rights, an incoherent mess. Nobody has the authority to enforce Emily's rights because nobody can, with a unilateral choice, place others under obligations that give themselves an entitlement to coerce (unilateralism). Nobody can make authoritative determinations of most of Emily's rights within the state of nature because there is nothing about them, naturally, that privileges their judgment over the judgment of others (indeterminacy). And nobody is under an obligation to respect others' rights to things they acquire—not only status rights but also property and contract rights—unless they have assurance that everyone will respect their rights to things (assurance). The result is that Emily's various rights are unenforceable because they are too indeterminate, nobody has the authority to determine them, and compliance with such a determination without assurance that others will comply is incompatible with our rightful honor.

Whether these defects arise does not depend on empirical facts about how people behave in the state of nature.[41] They arise, in Kant's view, "no matter how good and right-loving human beings might be."[42] For example, we do not need to wait around to see whether John and Alex disagree over Emily's education; indeterminacy over that right arises simply from the fact that it is *a priori disputable*, not that it is, in fact, disputed.[43] Or even if nobody disagrees with anything that Alex does for Emily because he is a perfectly virtuous parent, the defects of the state of nature still make the relationship incoherent from a matter of rights. As long as Alex claims a status right over Emily to clothe, feed, and educate her, he unilaterally asserts an entitlement to coerce others that he does not have the authority to enforce, even if no one challenges him.

1.3. Clarifications: Intervention, Provisional Right, and Children

Before moving on to the state's role in remedying this situation, let me make three clarifications. First, I do not mean to say John could *never* intervene in the parent-child relationship, for some of Emily's rights are determinate even in the state of nature. While Kantians are unique in insisting that most of our

41 For an opposing argument that says these defects are not necessary, see Christmas, "Against Kantian Statism."

42 Kant, *Metaphysics of Morals*, 6:312.

43 "Kant shows that rights are necessarily subject to dispute, not that they are always disputed. The application of concepts to particulars is always potentially indeterminate and so requires judgment, as a result of which the classification of particulars is always, at least in principle, indeterminate" (Ripstein, *Force and Freedom*, 170).

rights are too indeterminate within the state of nature, as opposed to Lockeans, who think nearly all our rights *are* determinate, albeit difficult to enforce, those relating to our bodies are not.[44] So suppose John sees Alex trying to murder Emily. While many of Emily's rights are indeterminate, her innate right determinately protects her body, so much so that there could be no possible reasonable disagreement over whether Alex is violating any of Emily's rights by trying to murder her. So if John intervenes to prevent Alex from murdering Emily, for instance, John is not standing on his authority as a unilateral enforcer of laws he has unilaterally determined; instead, he is simply standing on the natural authority of each person's innate right.[45] The salient difference between this case and the one above is that this case concerns the violation of a right that is already naturally determined. In contrast, the latter case concerns a right that is not naturally determined. So from the perspective of rights, the former intervention is permissible.

Moreover, there may be cases of intervention that are permissible as a matter of *virtue* but impermissible as a matter of *right*. For example, consider the following case.

> *John's Kidnapping*: John notices that Alex is a bad parent. For example, Alex gives his children just enough food to barely survive, does little for them when they become sick, socially isolates them so that they suffer severe psychological distress, and makes them do severe menial labor most of their waking moments. John decides to take Alex's children and care for them in the way that good parents are supposed to.[46]

What should we make of Alex's actions here? As a matter of virtue, maybe John did a good and praiseworthy thing (though even this seems disputable); however, as a matter of *rights*, John did the wrong thing. Perhaps John's kidnapping reveals that he is a benevolent and loving person who is properly concerned

44 However, for a compelling argument that even some of our bodily rights are indeterminate, see Pallikkathayil, "Persons and Bodies."

45 Pallikkathayil makes a similar point: "If the permissive law were itself justified by the requirements of freedom, the requirements of freedom would confer this authority on me. And . . . Kant does indeed regard equal freedom as requiring the possibility of ownership. So, I do not need to stand on my own authority [because] I can stand on the authority of the requirements of reason" ("Persons and Bodies," 45–46).

46 Some republicans may deny that John is dominating, which is why I should stress that I am defending only one kind of republicanism. Even with this qualification, though, I think other republicans will share my evaluation. For example, Henry Richardson says power is arbitrary when it fails to abide by a set of fair procedures "supported by independent, liberal ideals of respecting citizens as free and equal" (*Democratic Autonomy*, 38), which is close to my view. Thanks to Anne Jeffrey for pointing this out to me.

with the welfare of children. Still, from the perspective of *right*, John acted wrongly in multiple ways.

Even though many of Alex's children's rights are indeterminate in John's Kidnapping, John acts as if he has the natural authority to act like a legitimate state. By intervening, John unilaterally makes himself into a legislator by determining the rights of the children, the judge by deciding that Alex has violated his legislative determinations, and the enforcer by coercively taking Alex's children away from him. These are actions that only a public state can rightfully perform. Given the defects of the state of nature, no private individual can determine, enforce, and arbitrate previously undetermined rights without dominating. So heroic as John may be, he dominates by acting this way unilaterally, and Alex is entitled to resist John's attempt. The children may no longer be under a negligent parent, but they are so only because someone acted wrongly by kidnapping them. And worse, John's kidnapping simply restarts the problem of the children's domination. After all, John now claims a status right over the children that inherits all the *a priori* defects that Alex's original status right incurred. Such is the incoherence of the state of nature: one may act virtuously and wrongly at the same time.[47]

Second, when I denounce the parent-child relationship as slave-like in the state of nature, I do not mean to suggest that children should therefore run away from their parents. After all, children still need parental care if they are to develop into independent adults, and parents still have status rights over their children in the state of nature. However, in this condition, parental status rights are only *provisional* and inconclusive.[48] While Kant's doctrine of provisional rights is notoriously difficult, let me briefly sketch how provisional status rights work in general and then apply the view to the case of the family. If I have a provisional status right over my child, then I have "permission to coercively force anyone who does or might oppose [my] claim [over my child] into joining [me] in the civil society."[49] The provisional status of this right makes it importantly incomplete; it is not protected by a law "laying on each a duty of right not to make use of [my child] without your permission," and I lack the right to try and enforce such a law myself.[50]

The only way my status right becomes conclusive—i.e., one that is legitimately enforceable and one that everyone has an obligation to avoid interfering with on pain of my forcing you to refrain—is if we submit to a public state that can enforce, determine, and adjudicate this right. Because you are in

47 Ripstein, *Force and Freedom*, 160.

48 Ripstein, *Force and Freedom*, 165.

49 Ebels-Duggan, "Moral Community," 6. See also Kant, *Metaphysics of Morals*, 6:246.

50 Ebels-Duggan, "Moral Community," 4.

a community with me, and I have a provisional status right over my child, I can force you under a public state so that all our rights can finally become determined by a state that represents our wills. But until this happens, my provisional status right is insecure; I can assert it over my child, but I have to defend this right against those who would try to contravene it in the state of nature.[51] Once the public state is formed, my provisional status right over my child gives me a strong and presumptive claim against the state to conclude that right over my child.[52] All of this means, then, that parents can still claim status rights over their children in the state of nature, and even though these rights are merely provisional, they provide reason for children to stay with their parents nonetheless.

Third, some might wonder whether it makes sense to think of children as slaves who can be dominated because they do not yet have rational wills. Whereas a chattel slave owner who orders around another fully developed adult is clearly dominating the enslaved adult who should otherwise be able to direct their own life, a young child does not yet have the requisite agency needed to direct their own life. The idea that parents necessarily dominate newborns in the state of nature might be a category error; something has gone wrong, but surely it has nothing to do with the totally underdeveloped wills of the infants in question. In reply, I contend that babies can suffer domination in the state of nature, even if they have underdeveloped agency.[53] The reason is that babies, in virtue of being human, have an innate right to external freedom even if they do not yet have the capacity to exercise that right. In Joel Feinberg's terms, we can call this right an *innate right-in-trust*: it is a child's right against their parents to help the child develop into the sort of person who can eventually exercise the innate right that the child currently possesses only in a latent form.[54] From this right, we derive a formula: if a child needs *X* from their parents to develop into the sort of person who can eventually exercise their innate right, then the parents are required to provide *X* for the child. This sets up the fundamental problem of stateless families: if a child has no mechanism for the enforcement of these rights against their parents, the child suffers a dominating, slave-like relationship with their parents.

51 Kant, *Metaphysics of Morals*, 6:257.

52 Ebels-Duggan, "Moral Community," 7.

53 Some republicans may think that the right against domination is grounded in the possession of rational capacities, which babies lack. However, Kantian domination occurs when there is a violation of someone's innate right, a right we possess *in virtue of our humanity*, not in virtue of our possessing developed rational capacities. See Kant, *Metaphysics of Morals*, 6:281; and section 1.1 of this article. Other republicans also think babies have a right against domination that is not grounded in their rational capacities (e.g., Brooks, "Republican Children," 40–41, 52, 57).

54 See Feinberg, "The Child's Right to an Open Future."

2. THE SOLUTION TO THE SLAVE PROBLEM: THE STATE

I have contended that children are necessarily in a slave-like relationship with their parents in the state of nature. As long as children have a set of rights against their parents that no other third parties can enforce on their behalf, they are necessarily in dominating relationships that everyone else in the community has a duty to remedy. The only way the community can remedy this relationship is by submitting to a public state that can legislate, enforce, and adjudicate laws regulating the family.[55] It follows that the community must submit to such a state. Here, I briefly discuss how the state remedies this problem and why it is the only context in which the slave-like parent-child relationship can be remedied.

2.1. How the Three Branches Solve the Defects in the State of Nature

Recall the three problems in the state of nature: unilateralism, indeterminacy, and assurance. If a community establishes and submits to a public state that represents the community's will *omnilaterally* and not a particular individual's will *unilaterally*, all three problems are solved in one fell swoop through the ideal state's tripartite functions: (1) the executive branch solves assurance by providing a system of equal enforcement of the law, to which all citizens are held accountable; (2) the legislative branch solves unilateralism by making determinate laws that bind the community (as opposed to leaving individuals to stand on their authority to determine laws); and (3) the judicial branch solves indeterminacy by settling disputes over particular laws.[56] We can take these general state functions and apply them to the case of the parent-child relationship to get the following particular functions: (1) through its legislative function, the state codifies the normative network of the parent-child relationship in family law; (2) in its judicial function, the state arbitrates disputes about how family laws apply in specific cases; and (3) and in its executive function, the state enforces family law by prosecuting guilty parties and compensating victimized parties for their suffering. Once families exist under a state like this, the parental status rights involved are conclusive, and children are no longer subjected to slave-like relationships with their parents.[57]

Beyond these details, Kant does not give much more information on what the ideal state would look like. I think this is for good reason; after all, ideal constitutions manifest differently depending on the nature of the people they govern. For example, the conditions for representing a unilateral will in a

55 Varden, "A Kantian Critique of the Care Tradition," 340.

56 Ebels-Duggan, "Kant's Political Philosophy," 899.

57 Varden, "A Kantian Critique of the Care Tradition," 340–41.

government of merely ten people can plausibly be much more stringent than the conditions that represent three hundred million people. Nonetheless, in section 3, I explore two possible policies that reflect the state's role in safeguarding children's rights: parental licensing and monitoring. In short, I argue that there are no good in-principle objections against either of these proposals to the effect that legitimate states lack the authority to enact either policy. On the contrary, the only reason states should not enforce either of these policies is *prudential*—i.e., some empirical fact makes these policies inept at protecting children without domination. Of course, this does not mean that licensing and monitoring are often permissible. Contingent empirical facts might always make these policies inept at achieving their goals. Nonetheless, it is important to underscore that these empirical considerations, not *a priori* objections that these policies are not within the state's purview to enforce, are ultimately what make or break the permissibility of licensing and monitoring.

Before discussing these policies in section 3, though, we should consider the state's role in creating *legal* parent-child relationships. Once we understand this, we can evaluate why the state is required to ensure that children are placed in the custody of *competent* parents, given the civil liability that the state incurs when it creates any given legal parent-child relationship. After this discussion, we can consider how they bear on licensing and monitoring.

2.2. Civil Liability and the Creation of Legal Families

It is tempting to believe that the state intervenes in the parent-child relationship only in severe, dramatic cases of abuse and neglect. However, this is not the case. On the contrary, *the state is responsible for concluding every parent-child relationship under its domain*, which means every legal parent-child relationship exists in virtue of some positive state action. As James G. Dwyer argues,

> The state first creates legal parent-child relationships (who else could create a *legal* relationship?), and for that legal relationship to have practical significance, the state must also confer on the persons whom it has made legal parents some particular privileges and powers. . . . People occupy legally protected custodial and caretaking roles as to children because the state places them in that role. This is as true of biological parents as it is of foster parents and adoptive parents. . . . A child's first legal parent-child relationship generally arises by virtue of state statutes that operate without any court involvement . . . so the state action is unrecognized by most people, but it exists nonetheless.[58]

58 Dwyer, "Regulating Child Rearing in a Culturally Diverse Society," 359–60.

No matter the kind of parent-child relationship, the state always intervenes through the law to create a legal, custodial parent-child relationship. Even in the most standard cases where two procreators have a child and are ideally suited to meet all the child's interests, there is a background state law that affords them custodial rights when they sign the birth certificate at the hospital.

Of course, this does not mean that the parent-child relationship is purely a legal convention. In section 1.3, I argued that the role of the public state is to conclude the provisional status rights that parents have over their children before any positive state action is taken. The provisional status right exists independently of any state action. For example, if a couple conceives a child in the state of nature, their status right over their child is provisional and deeply insecure as long as there is no public state available that can conclude that right. But if a couple conceives a child under a public state, they have a provisional status right over their child that the state should conclude. Notably, this provisional status right is a strong *presumptive* claim against the state to conclude it in the form of a legally protected parent-child relationship, though the state can deny this claim in some cases—i.e., when concluding the claim would violate the prohibition against care uncertainty. So there is a sense in which the state has relatively wide discretion in which positive laws it can enact; however, that discretion is circumscribed by any strong provisional rights its citizenry possesses.

Now, if the state is responsible for the existence of every legal parent-child relationship, then it follows that the state is liable should it place a child with sufficiently incompetent parents.[59] This principle of civil liability explains other cases in which the state is responsible for things it creates. For example, suppose the state builds a bridge that collapses a few years later because it was poorly constructed. Any victims of the collapse have cause to sue the state because it is civilly liable for the collapse. It created the bridge for public transportation, and because its negligent construction caused damage to those using it, it owes the victims some form of compensation.[60]

Similarly, if the state places a child under the custody of sufficiently incompetent parents without making a reasonable effort to ensure the child's parents

59 Similarly, Dwyer claims the state is importantly responsible for violations of children's rights when it consigns them to unfit parents. See Dwyer, "Child Protection Pretense," pt. 1.

60 In the United States, the Federal Tort Claims Act describes such liability. Under this edict, "the federal government acts as a self-insurer and recognizes liability for the negligent or wrongful acts or omissions of its employees acting within the scope of their official duties. The United States is liable to the same extent an individual would be in like circumstances." See the House of Representatives webpage "Federal Tort Claims Act," https://www.house.gov/doing-business-with-the-house/leases/federal-tort-claims-act (accessed November 6, 2025).

are competent enough for the job, then it is liable to the child and owes the child compensation for its failure. After all, the only reason the state has the authority to create the custodial relationship in the first place is that it is playing the advocate role for the child that the child would otherwise lack in the state of nature. If the state puts together custodial parent-child relationships without trying to ensure the child's various rights will be taken care of, it fails its duty to advocate for the child's rights against the parents. Worse than this, in virtue of subjecting the child to a legal custodial relationship with an incompetent parent, it also becomes partially responsible for rights violations the child suffers through parental incompetence.[61]

This account of the state-family relationship is quite different from prepolitical accounts that suggest the family can exist nondefectively before any state action. For example, Melissa Moschella claims that "the family [is] a prepolitical authoritative community."[62] As a result, she likens the threshold for state intervention in the family to that of neighboring countries. Just as states—entities that exist independently of each other—should not intervene on one another unless they get sufficient evidence of significant human rights abuses, so too should states refrain from interfering with families—another entity that Moschella thinks can exist rightfully, independent of the state—unless the state receives evidence of the same kind of abuses (though not necessarily the same scale).[63] Or consider philosophical anarchists, who claim that states generally have no right to intervene in a citizen's life unless the citizen consents to such intervention or unless the state is fulfilling a sufficiently strong duty of rescue.[64] In light of the state's role in safeguarding the parent-child relationship and the liability it incurs in virtue of creating every legal parent-child relationship, neither of the aforementioned

61 Dwyer, "No Place for Children," 924–25.

62 Moschella, *To Whom Do Children Belong?* 45.

63 Moschella, *To Whom Do Children Belong?* 67–68. In fairness to Moschella, she clarifies that "the family's self-sufficiency with respect to the larger political community is much more limited than that of the state with respect to the larger international community," which means that "the abuse or threat to the public order [in the family] need not be as serious or immediate to justify state intervention into the family sphere as to justify international intervention" (68). Nonetheless, it is hard to avoid the worry that her view requires the extremely strong presumption against interference in the international case to govern the state-family case if the family really is its own authoritative body that can exist nondefectively, prepolitically. This worry seems especially strong since Moschella thinks that "the grounds justifying coercive intervention" in the international and state-family cases—namely, human rights and protection of public order—"remain essentially the same," even after our starting qualification (*To Whom Do Children Belong?* 68). For more on her account, see pp. 66–71.

64 For example, see Huemer, *The Problem of Political Authority*.

views is correct. Moschella's foreign policy analogy is too bare because it overlooks the state's role in concluding the parent-child relationship and in keeping children from slave-like relationships. Whereas neighboring countries do not necessarily incur any liability for another country's disorder, states are uniquely liable to children because they are the ones who consign children to particular familial arrangements. And if the state has a Kantian role in safeguarding children's rights against their parents, then protests from philosophical anarchists that parents have not consented to state intervention or that the state's intervention is not required to meet some strong duty of rescue are immaterial.

3. Evaluating Licensing and Monitoring

3.1. Licensing

Parental licensing is a policy whereby "the state [uses] evaluative tools to determine whether individuals are competent to be parents before raising children and excluding those who are judged to be incompetent from raising children."[65] Proponents of this policy have views about which evaluative tools the state should use, why the state should enforce them, and what should happen when an applicant fails a licensing test. (For example, extreme versions of parental licensing may entail forced separation as a result of failure, while mild versions may entail missing out on beneficial tax breaks.)[66] I do not arbitrate these debates here; I just want to show how *some* licensing schemes using *some* standards can help states safeguard children's rights under the appropriate circumstances. I start with a plausible licensing scheme in the United States and then speculate about other constitutions.

The content of a child's rights determines the state's licensing standards. If children have rights only against abuse and neglect, then the standards are minimal. Maybe all that parents need to do in such a case is fill out a form to let the state know they have not had prior terminations of parental rights. Even though this query is so minimal that it seems to stretch the meaning of a licensing evaluation, it is already several steps above the American status quo, whereby the state treats these facts as immaterial to whether it grants procreators custodial rights over their offspring.[67] In the United States, procreators face no legal barrier to acquiring custodial rights over their children so long as they are not currently in prison—a temporary obstacle that they can overcome once their sentences are over—or, in particular jurisdictions, so long as they are not guilty of sexual

65 Kianpour, "The Kids Aren't Alright," 432.

66 LaFollette, "Parental Licensing Revisited," 338–40.

67 Dwyer, *Liberal Child Welfare and the Destruction of Black Lives*, 66.

assault against the mother.[68] So one step the United States could take toward protecting children is simply taking into consideration the fact that a procreator has lost their parental rights in the past before reissuing custodial rights.[69]

However, children's rights seem to go beyond a mere right against abuse and neglect. On the normative network of the parent-child relationship described above, children have the right to whatever they need to become independent agents. While this means they have a right against abuse and neglect, they also need other protections to ensure their proper development. For example, children require a minimally adequate education that will equip them to determine which conception of the good life they wish to pursue when they get older.[70] They also appear to require parental love, which seems vital to their proper development.[71] Here, I do not give a full account of everything children have a right to; instead, I suggest a formal apparatus: if children generally have a right against their parents to become independent agents, then they have a right to those things they need to become independent agents. If the goal of parental licensing is to ensure that parents are competent to meet their children's various rights, then whatever licensing standards the state uses should be sensitive to these rights. For example, if children need parental love in order to become independent persons, then it is within the state's authority to use standards that effectively measure whether parents can in fact meet that goal (if such standards exist).

Some might object that this scheme would not work and that it is too implausible given how many parents and children there are. This may very well be true, which is why I defend parental licensing as, *in principle*, permissible for the state to enact. But if it turns out that in practice, such a scheme is impossible to permissibly enforce given facts on the ground, then it should not be enforced. However, notice that this is a *contingent* worry; there may be some states that are small enough to effectively license all parent-child relationships with all relevant

68 Dwyer, *Liberal Child Welfare and the Destruction of Black Lives*, 66.

69 Some worry that punishing someone for past terminations of parental rights is unjustly disproportionate. But we can quell this worry in three ways. First, some licensing schemes do not involve denying anyone custodial rights—e.g., see a recent proposal by LaFollette that licenses could simply get parents tax breaks ("Parental Licensing Revisited," 338–40). Second, licensing is not necessarily a punitive mechanism (Baron and Cowley, *Philosophy of the Family*, 21). For example, the state is not punishing someone who fails a driving test when it refuses them a driver's license. Third, even on licensing schemes that deny custodial rights—which I take no stand on here—it is possible that applicants could reapply.

70 Brighouse and Swift, *Family Values*.

71 Liao, *The Right to Be Loved*; and Ferracioli, "The State's Duty to Ensure Children Are Loved." In the latter paper, though, Ferracioli sidesteps the empirical debate about whether children need love for proper development.

standards. In other states, licensing schemes may be enforceable because the licensing state has more resources or fewer applicants. However, in larger states, the licensing state may have to either reduce the level of involvement in the standards or relinquish its aspirations to license altogether. So whether a particular state should enforce a licensing scheme is an empirical matter that cannot be settled here. The important point is that if empirical facts on the ground support a licensing scheme that would help the state ensure parental competence without domination, the state can permissibly enforce such a policy.

3.2. Monitoring

The same rationale that justifies parental licensing—i.e., that states are required to advocate for a child's rights within the family—also justifies parental monitoring.[72] In short, parental monitoring is a policy whereby representatives of the state visit households to assess the welfare of the children within them.[73] If the state's advocate role justifies licensing, it also justifies monitoring since parents can plausibly become unfit to raise a child even after passing the licensing test at the beginning of their custodial relationship. As with parental licensing, proponents of this policy do not share a single, unified vision for how it should be implemented. However, there is generally a sense that the frequency of monitoring should decrease as a child develops into someone who can advocate for themselves.

Still, this would be a marked departure from the status quo. At least in the United States, child protective services work on a wait-and-see basis, and intervention can occur only when sufficient evidence of sufficiently extreme parental abuse and neglect comes to the state's attention.[74] Organizations like child protective services "are safeguards only against the very worst forms of abuse and neglect, and they are highly fallible safeguards."[75] As Hugh LaFollette notes, American child protective services process millions of substantiated cases of abuse and neglect every year, and there is good reason to think the real number of abuse and neglect is several times higher than that.[76] The status quo wait-and-see approach hamstrings social workers' abilities to respond effectively to evidence, which also results in many unjust interventions among Black and working-class families.[77] Moreover, status quo American family law affords

72 Archard, *Children*, 190–91.

73 Kianpour, "The Kids Aren't Alright," 432.

74 Archard, *Children*, 173.

75 Gheaus, "Children's Vulnerability and Legitimate Authority over Children," 62.

76 LaFollette, "Parental Licensing Revisited," 336.

77 Roberts, *Torn Apart*.

parents such strong rights over their children that this problem is currently intractable. There is no legal requirement for parents to ensure that their children have any relational contact with third parties such as relatives, neighbors, or close family friends. This issue is especially acute for preschool-aged and homeschooled children, whose parents have the legal power to prevent third parties from associating with their children at all.[78]

One of the main reasons that parents have such strong rights over their children is presumably a strong liberal emphasis on the right to privacy.[79] But we know that too much privacy has the power to enable and hide parental violations of children's rights, as David Archard notes:

> The ill-treatment of children takes place in a "private" space, the family home, and to the very extent that it is a private space, it may continue undetected and unsuspected. . . . The abuse is literally unobserved, and whilst physical abuse may display itself . . . sexual abuse has no obvious public face. Abused children may have no sense of what is privately happening to them is radically and terribly different from what would be publicly acceptable. . . . Finally, the abused child within the private space of the family will probably be pressurized not to reveal the abuse, or retract previous accusations of abuse.[80]

None of this is to say that parents lack a right to privacy; they certainly have one, especially if privacy is necessary to procure some important goods that children need for their proper development. However, the type of intervention that monitoring involves—an occasional visit from a qualified state representative to a family every year or so—does not seem to infringe upon that right and therefore does not seem to jeopardize the welfare of the children in question.

So what would a monitoring policy look like? As mentioned, whether and how states enforce such a policy is a highly complex matter that I do not pretend to settle from the armchair. Nonetheless, we can make a few very tentative suggestions for what enforcement might look like.

1. A representative from the state could visit families annually for the first few years of a child's life and gradually reduce the frequency as the child gets older.
2. If there are too many families to monitor on an annual basis, families could face audits similar to those that the Internal Revenue Service carries out on taxes. For example, families could be visited by

78 Gheaus, "Children's Vulnerability and Legitimate Authority over Children," 63.

79 For example, see Fried, "Privacy."

80 Archard, "Child Abuse," 192.

representatives through randomized lotteries based on family census data. If a family is selected for an audit, they could be given short notice to return home and make time for a representative to visit them and evaluate the child(ren).

3. Parents who homeschool their children may be required to make sure a representative of the state has some semiregular contact with their children as they get older. For example, if there is a homeschooling co-op, perhaps a teacher who is suitably trained as a mandatory reporter could act in a public role and submit evaluations on the students they see.

Of course, these are just speculative ideas, and each state should make prudential decisions about which ones to enact or reform. But the point is that monitoring takes seriously the fact that children are extremely vulnerable to their parents for most of their childhoods, and the state advocates for children's rights not merely at birth but also throughout the entirety of the time that children are under their parents' custodial authority. If empirical facts allow for a monitoring scheme to protect children without domination, states may rightfully enforce it.

4. Conclusion

In sum, I have argued that children need states to avoid slave-like relationships with their families, and this need gives the state the normative basis for stepping in and advocating for children's rights through family law. I have also argued that from this foundation, there are no good in-principle objections to child welfare schemes like parental licensing and parental monitoring. Of course, this does not mean that all states should enforce these policies. Every state must account for empirical conditions that make the enforcement of either policy more or less prudent. But given the state's relationship to the family, we should oppose these policies only for prudential reasons, not by appealing to arguments that such policies are beyond the state's authority to enforce.[81]

West Virginia University
nicholas.hadsell@mail.wvu.edu

81 Thanks to Anne Jeffrey, Kyla Ebels-Duggan, and Matthew Lee Anderson for helping me to think through the main moves of this article. Thanks also to Alexander Pruss, C. Stephen Evans, Francis J. Beckwith, the journal's anonymous reviewers, and the audience at the 2025 Central Division meeting of the American Philosophical Association for helpful feedback.

REFERENCES

Allais, Lucy. "What Properly Belongs to Me." *Journal of Moral Philosophy* 11, no. 4 (2014): 754–71.

Archard, David. "Child Abuse: Parental Rights and the Interests of the Child." *Journal of Applied Philosophy* 7, no. 2 (1990): 183–94.

———. *Children: Rights and Childhood*. 2nd ed. Routledge, 2004.

———. "The Obligations and Responsibilities of Parenthood." In *Procreation and Parenthood: The Ethics of Bearing and Rearing Children*, edited by David Archard and David Benatar. Oxford University Press, 2010.

Baron, Teresa, and Christopher Cowley. *Philosophy of the Family: Ethics, Identity and Responsibility*. Bloomsbury Academic, 2024.

Brighouse, Harry, and Adam Swift. *Family Values: The Ethics of Parent-Child Relationships*. Princeton University Press, 2014.

Brooks, Thom. "Republican Children." *Philosophy and Public Affairs* 53, no. 1 (2025): 37–65.

Christmas, Billy. "Against Kantian Statism." *Journal of Politics* 83, no. 4 (2021): 1721–33.

Clayton, Matthew. *Justice and Legitimacy in Upbringing*. Oxford University Press, 2006.

De Wispelaere, Jurgen, and Daniel Weinstock. "Licensing Parents to Protect Our Children?" *Ethics and Social Welfare* 6, no. 2 (2012): 195–205.

Dwyer, James G. "The Child Protection Pretense: States' Continued Consignment of Newborn Babies to Unfit Parents." *Minnesota Law Review* 93 (2008): 407–92.

———. *Liberal Child Welfare Policy and Its Destruction of Black Lives*. Routledge, 2018.

———. "No Place for Children: Addressing Urban Blight and Its Impact on Children Through Child Protection Law, Domestic Relations Law, and 'Adult-Only' Residential Zoning." *Alabama Law Review* 62, no. 5 (2012): 887–961.

———. "Regulating Child Rearing in a Culturally Diverse Society." In *Philosophical Foundations of Children's and Family Law*, edited by Elizabeth Brake and Lucinda Ferguson. Oxford University Press, 2018.

Ebels-Duggan, Kyla. "Kant's Political Philosophy." *Philosophy Compass* 7, no. 12 (2012): 896–909.

———. "Moral Community: Escaping the Ethical State of Nature." *Philosopher's Imprint* 9, no. 8 (2009): 1–19.

Feinberg, Joel. "The Child's Right to an Open Future." In *Justice, Politics, and the Family*, edited by Daniel Engster and Tamara Metz. Routledge, 2014.

Ferracioli, Luara. "Citizenship for Children: By Soil, by Blood, or by Paternalism?" *Philosophical Studies* 175, no. 11 (2018): 2859–77.
———. *Parenting and the Goods of Childhood*. Oxford University Press, 2023.
———. "The State's Duty to Ensure Children Are Loved." *Journal of Ethics and Social Philosophy* 8, no. 2 (2014): 1–20.
Freiman, Christopher. "Against Parental Licensing." *Journal of Social Philosophy* 53, no. 1 (2022): 113–26.
Fried, Charles. "Privacy." *Yale Law Journal* 77, no. 3 (1968): 475–93.
Gheaus, Anca. "Children's Vulnerability and Legitimate Authority over Children." *Journal of Applied Philosophy* 35, no. S1 (2018): 60–75.
Hadsell, Nicholas. "Kant on Punishment and Poverty." *Southern Journal of Philosophy* 62, no. 2 (2024): 193–210.
Huemer, Michael. *The Problem of Political Authority: An Examination of the Right to Coerce and the Duty to Obey*. Palgrave Macmillan, 2012.
Kant, Immanuel. *The Metaphysics of Morals*. Edited by Lara Denis, translated by Mary Gregor. Cambridge University Press, 2017.
Kianpour, Connor K. "The Kids Aren't Alright: Expanding the Role of the State in Parenting." *Journal of Ethics and Social Philosophy* 25, no. 3 (2023): 431–63.
Kolodny, Niko. "Being Under the Power of Others." In *Republicanism and the Future of Democracy*, edited by Geneviève Rousselière and Yiftah Elazar. Cambridge University Press, 2019.
LaFollette, Hugh. "Licensing Parents Revisited." *Journal of Applied Philosophy* 27, no. 4 (2010): 327–43.
Liao, S. Matthew. *The Right to Be Loved*. Oxford University Press, 2015.
Locke, John. *Two Treatises of Government*. Edited by Peter Laslett. Cambridge University Press, 1988.
Loriaux, Sylvie. "Kant on Social Justice: Poverty, Dependence, and Depersonification." *Southern Journal of Philosophy* 61, no. 1 (2023): 233–56.
Moschella, Melissa. *To Whom Do Children Belong? Parental Rights, Civic Education, and Children's Autonomy*. Cambridge University Press, 2016.
Neufeld, Blain. "Coercion, the Basic Structure, and the Family." *Journal of Social Philosophy* 40, no. 1 (2009): 37–54.
Pallikkathayil, Japa. "Deriving Morality from Politics: Rethinking the Formula of Humanity." *Ethics* 121, no. 1 (2010): 116–47.
———. "Persons and Bodies." In *Freedom and Force: Essays on Kant's Legal Philosophy*, edited by Sari Kisilevsky and Martin Jay Stone. Bloomsbury Academic, 2017.
Porter, Lindsey. "Why and How to Prefer a Causal Account of Parenthood." *Journal of Social Philosophy* 45, no. 2 (2014): 182–202.
Richardson, Henry. *Democratic Autonomy: Public Reasoning About the Ends of*

Policy. Oxford University Press, 2003.
Ripstein, Arthur. *Force and Freedom: Kant's Legal and Political Philosophy*. Harvard University Press, 2009.
Roberts, Dorothy. *Torn Apart: How the Child Welfare System Destroys Black Families—and How Abolition Can Build a Safer World*. Basic Books, 2022.
Shelby, Tommy. *Dark Ghettos: Injustice, Dissent, and Reform*. Harvard University Press, 2016.
Simmons, A. John. *The Lockean Theory of Rights*. Princeton University Press, 1994.
Smith, Margaret, and Rowena Fong. *Children of Neglect: When No One Cares*. Routledge, 2003.
Stilz, Anna. *Liberal Loyalty: Freedom, Obligation, and the State.* Princeton University Press, 2009.
Stone, Martin Jay, and Rafeeq Hasan. "What Is Provisional Right?" *Philosophical Review* 131, no. 1 (2022):51–98.
Varden, Helga. "A Kantian Critique of the Care Tradition: Family Law and Systemic Justice." *Kantian Review* 17, no. 2 (2012): 327–56.
———. "Kant's Non-Voluntarist Conception of Political Obligations: Why Justice Is Impossible in the State of Nature." *Kantian Review* 13, no. 2 (2008): 1–45.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.3501

GRATITUDE'S FITTING GROWTH

Daniel Telech

GRATITUDE is a response not simply to being benefited but to being benefited from good will—that is, to being *benevolently* benefited. Natural events and malicious persons may cause some end of mine to be fulfilled, and while I may resultantly be glad that that beneficial state of affairs arose, I am not ordinarily thereby disposed to feel grateful to or thank anyone. This is because gratitude—at least that of the central sense of interest to me—is a response *to* another *for* a benevolently given benefit.[1] Peter Strawson expresses this thought about gratitude by distinguishing between the response to merely being benefited, on the one hand, and the response to being benefited *from good will*, on the other: "If someone's actions help me to some benefit I desire, then I am benefited in any case; but if he intended them so as to benefit me because of his general good will towards me, I shall reasonably feel gratitude which I should not feel at all if the benefit was an incidental consequence unintended or even regretted by him, or some plan or action of a different aim."[2] On this point, Strawson is in good company; the claim that gratitude construes another to have manifested good will is a mainstay in philosophical discussion of gratitude.[3]

Little discussion, however, has been devoted to the responsiveness of gratitude to *outcomes*—particularly the unforeseen and unintended outcomes—of manifestations of good will. Although the above picture of gratitude does not preclude gratitude's being appropriately outcome sensitive, given its focus on

1 Another sense of 'gratitude', employed in sentences like 'The farmer is *grateful that* it rained', fails to pick out an inherently interpersonal emotion. This other reaction, *propositional gratitude*, involves a dyadic relation between a person and state of affairs (or proposition), with no essential reference to benevolent agency. See McAleer, "Propositional Gratitude"; and Rush, "Motivating Propositional Gratitude." This is not my topic. By 'gratitude', I mean the essentially agent-directed emotion involving a triadic relation between two agents and (typically) an action, as in 'Abe is grateful *to* Miranda *for* helping him move'.

2 Strawson, "Freedom and Resentment," 62.

3 Berger, "Gratitude," 299–300; Walker, "Gratitude and Gratefulness"; Camenisch, "Gift and Gratitude in Ethics"; McConnell, *Gratitude*; Roberts, "The Blessings of Gratitude"; Manela, "Gratitude and Appreciation"; Roberts and Telech, "The Emotion-Virtue-Duty Triad of Gratitude"; and Macnamara, "The Emotion of Gratitude and Communal Relationships."

benevolence, it is unsurprising that the question has gone largely unaddressed.[4] This article investigates the possibility that gratitude may be appropriately outcome sensitive by reflecting on cases in which gratitude is amplified owing to the unforeseen consequences of a manifestation of good will. To anticipate, I propose that gratitude can be appropriately outcome sensitive, but this requires more than a causal relation between the original manifestation of good will and the fortunate outcome.

To clearly distinguish outcomes of manifestations of good will from those effects of a benefactor's intentions that are part of the manifestation of good will (e.g., the intended benefit), it is useful to restrict our discussion to cases in which the gratitude-amplifying outcomes occur long after the initial manifestation of benevolence.[5] In such cases, the outcomes are not only unforeseen and unintended (which can be stipulated in any case); it is also implausible that the benefactor could have foreseen and intended such distant outcomes. By attending to cases of diachronic gratitude that span many years, we are provided a "macroscale" view of the target phenomenon.

Accordingly, consider the following case.

> *Musical Fulfillment*: When he is eleven years old, Tim receives a baroque flute from his aunt, Ida. Ida has taken Tim to several baroque performances in the past, which Tim has enjoyed. After some reflection on what instrument might suit him, Ida has decided on the flute. Tim is touched by his aunt's thoughtfulness, appreciates the gift, and expresses his gratitude with a heartfelt hug and words of thanks. Years pass without Tim's thinking of the flute. But on his seventeenth birthday, having watched Ingmar Bergman's film *The Magic Flute*, Tim decides to play with the instrument that his aunt gave him years ago. Finding the experience enjoyable, Tim begins taking flute lessons. To his surprise, he displays a knack for the instrument. After considerable time practicing, followed by a series of increasingly important recitals, Tim emerges as a virtuoso, eventually

4 This is the case at least in recent literature. Adam Smith addresses gratitude's sensitivity to features of benefits "beyond what is due to the motives or affections from which they proceed" (*The Theory of Moral Sentiments*, II.iii.2.1) and thus touches on the general issue of interest to me. But Smith maintains that gratitude's sensitivity to more than quality of will is warranted by its social utility. As I discuss below, the justification of interest to me is of a different kind from that afforded by forward-looking considerations, like those of utility.

5 So while it is possible to individuate actions such that any benefit is an outcome—i.e., an *outcome of the benefactor's intentions*—as I employ these terms, the intended benefit that the benefactor brings about is one thing (the manifestation of good will, or at a least a *part* of the manifestation of good will), and the downstream consequences of the benevolently given benefit are another.

> becoming a highly skilled and esteemed flautist. By the time he is in his mid-thirties, Tim is a world-renowned flautist whose craft (and "calling," as he refers to it) provides him with a sense of meaning in life that he never thought possible as a young man. He is immensely grateful to his aunt. Thinking of it as a token of his gratitude, he dedicates (and then mails several copies of) his first recorded album to her. The record sleeve reads, "To Aunt Ida, with gratitude, for introducing me to my calling —Tim."

Call the time that Tim receives the flute t_1. Call the time from which Tim takes being a flautist to be his calling and career t_2. At t_1, Tim is grateful to his aunt for gifting him a flute. At t_2, Tim is considerably *more grateful* to his aunt. That is, he is grateful *to a greater degree*.[6] Yet Tim is not and does not take himself to be the recipient of any greater degree of benevolence between t_1 and t_2.

Extrapolating from its details as presented, note that the case has the following structure: (1) a person (S) is grateful at some time (t_1) to another person (V) for ϕ-ing; (2) S is at some later time (t_2) *more grateful* to V for ϕ-ing, even though (3) no change in V's degree of benevolence has occurred between t_1 and t_2; and (4) S knows that 3. I use the term 'gratitude growth' to refer to the this phenomenon. Musical Fulfillment is meant to be an example of gratitude growth.

My primary question is: Can gratitude growth be fitting? To say that an instance of an emotion is fitting is to say that it accurately appraises that which it is about.[7] As emotions are assessable relative to standards other than fittingness—e.g., moral or prudential standards—there are other ways that emotions

6 In emphasizing that Tim's gratitude increases in *degree*, I mean to indicate that the change in question is not merely a change in the *scope* of his gratitude—i.e., what Tim's gratitude is about/what he is grateful for. See Zimmerman, "Taking Luck Seriously," 560. I clarify in section 2 below how I understand "more grateful," but to anticipate, I take this to involve an increase in *both* the appreciative and motivational elements of gratitude, which I introduce in section 1.

7 D'Arms and Jacobson, "The Moralistic Fallacy" and *Rational Sentimentalism*. I understand appraisal capaciously, as including not only the representational content of the emotion but also its motivational profile concerning its object. Fear of the wolf, for example, not only represents the wolf as dangerous—a *judgment* concerning the wolf could do this—but also (perhaps, thereby) poises one to flee, given the wolf's perceived dangerousness. The motivation to flee is not arational but something that is *justifiable* by reference to the considerations that render fear's representation accurate—namely, the wolf's dangerousness. The emotional motivation to flee can be excessive or deficient and, as such, unfitting even if one has a negatively valanced emotion that represents the wolf as dangerous. How exactly to understand the relation between an emotion's representational content and its motivational profile—including whether the fittingness of the latter can be reduced to that of the former—is beyond our scope here. For discussion, see Mason, "Dimensions of Emotional Fit." For a holistic view of emotional appraisal as including evaluation and motivation, see D'Arms and Jacobson, *Rational Sentimentalism*.

can be (in)appropriate.[8] But here, I am interested in gratitude's appropriateness *qua* fittingness. Consider what we mean in saying that your gratitude is a fitting response to, e.g., an acquaintance for kindly helping you with some distressing problem. Here, gratitude is fitting in that it accurately appraises the relevant evaluative qualities of the action—namely, its benevolent and beneficial qualities. Your acquaintance is, as we can put it, *gratitudeworthy* for benevolently helping you.

Can gratitude growth be fitting? It is not obvious how it can be. Gratitude is a backward-looking emotion; it is about and justified (when it is) by something past. Particularly, gratitude represents some past action as a benefit to oneself motivated by another's good will. It is fitting when it correctly represents another to have benevolently benefited oneself.[9] But given that the past is unalterable, gratitude growth is presumably unfitting. For if gratitude of a certain sort was a fitting response at t_1, on the basis of a manifestation of good will, ϕ, which occurred prior to t_1, given that ϕ's occurrence is fixed in the past, how can a greater degree of gratitude for ϕ be fitting at t_2? [10]

8 Gratitude is thought to promote subjective well-being and relationship quality. See Emmons and McCullough, "Counting Blessings Versus Burdens"; Seligman, *Authentic Happiness*, 74; Peterson and Seligman, *Character Strengths and Virtues*, 524; Wood et al., "Gratitude and Well-Being." To the extent that being *more* grateful contributes to one's well-being and relationships, there may be prudential and moral reasons in favor of gratitude growth. I return below to these kinds of reasons—i.e., reasons of the "wrong kind," which do not bear on whether the emotion accurately appraises its target and (for that reason) are not considerations on the basis of which we can *directly* feel emotions. That is, even if some evil demon's credible threat of "Admire me, or else I'll torture you and other innocents" is a good reason to (come to) admire him, being a wrong kind of reason, it does not render the demon admirable—i.e., a fitting target of admiration. It provides you not direct (i.e., fitting) reason for admiration but indirect reason to admire him—e.g., to act in a way that enables one to construe the demon as admirable. See D'Arms and Jacobson, "The Moralistic Fallacy"; Rabinowicz and Rønnow-Rasmussen, "The Strike of the Demon"; and Hieronymi, "The Wrong Kind of Reason."

9 Strictly speaking, this should be "oneself or one with whom one (correctly) *identifies*," for one can feel grateful to another without construing oneself as the beneficiary. See Roberts and Telech, "The Emotion-Virtue-Duty Triad of Gratitude," 1; and Lewis, "Gratitude," 2. For simplicity's sake, however, the grateful agents discussed in this article are beneficiaries.

10 Although I characterize gratitude growth as an increase in the *degree* to which gratitude is fitting, nothing hangs on the assumption that when an agent is gratitudeworthy, there is a single (i.e., unique) degree of gratitude that is fitting. It may be that whenever an agent is gratitudeworthy for some action, there is a bounded *range* of degrees of gratitude that is fitting. On this point as it concerns blameworthiness, see Sommers, "Partial Desert," 255–56; and Telech and Tierney, "The Comparative Nonarbitrariness Norm of Blame," 33–35. Those sympathetic to this view can read gratitude growth as a thesis concerning an increase in the range of degrees to which gratitude is fitting. For example, perhaps it was fitting at t_1 for Tim to feel gratitude toward Ida within a range (R_1) of degrees (D_1–D_5),

But although gratitude growth, considered in the abstract, may appear unfitting, when we reflect on particular cases, outcomes of manifestations of benevolence intuitively *do* sometimes render fitting a greater degree of gratitude than was fitting in response to the original manifestation of benevolence. Granting, as per Strawson and others, that beneficial outcomes cannot themselves render gratitude fitting—that is, in the absence of benevolence—it is puzzling why outcomes should sometimes appear to make a difference to gratitude's fittingness.

The puzzle concerning gratitude growth can be fruitfully viewed in the context of a broader discussion about the rationality of emotional change. Philosophers have recently attended to several backward-looking emotions that, despite the constancy of the facts that presumably render them fitting, appear to fittingly change across time.[11] For example, anger (or resentment) is a fitting response to wrongdoing, but as the wrongdoing cannot be undone, it is unclear how anger fittingly diminishes. Yet when we forswear anger (or resentment) in forgiving wrongdoers, the diminution of anger that often follows seems not only statistically ordinary but also fitting—i.e., it seems a correct response to the relevant evaluative phenomena. Similarly, if grief is a fitting response to the loss of a loved one, given that the loss cannot be *unlost*, how can grief fittingly subside? While different philosophers unsurprisingly offer different responses about different backward-looking emotions, the literature displays a common tendency—namely, a focus on the *diminishment* of emotion. Indeed, the problem concerning backward-looking emotions is sometimes framed as the problem of the diminution of backward-looking emotion.[12] In attending to the possible fittingness of gratitude's *strengthening* across time, this article makes an indirect case for broadening the scope of the discussion concerning changes in backward-looking emotions.

In what follows, I attend to the relationship between gratitude and factors external to quality of will, particularly the fortunate consequences of manifestations of benevolence. I focus on cases of diachronic gratitude, arguing that the (unforeseen and unintended) consequences of a benefactor's manifestation of

---

and at t_2, Ida is more gratitudeworthy, such that she is, from that time, the fitting target of a greater range (R_2) of degrees (D_{10}–D_{15}) of gratitude.

11 For example, Callard, "The Reason to Be Angry Forever"; Cholbi, "Grief's Rationality, Backward and Forward"; Hieronymi, "Articulating an Uncompromising Forgiveness"; Marušić, "Do Reasons Expire?"; Moller, "Love and Death"; Na'aman, "The Fitting Resolution of Anger"; and Schönherr, "Two Problems of Fitting Grief."

12 For example, Na'aman, "The Fitting Resolution of Anger." But see Na'aman, "The Rationality of Emotional Change," 260–61. Additionally, the literature focuses on changes to negatively valenced emotions. But for exceptions, see Archer and Matheson, "Admiration over Time"; and Coren, "Giving Up Gratitude."

good will can render fitting the beneficiary's being more grateful to her benefactor than she was for the initially willed benefit. It is only a subset of fortunate consequences, however, that can render fitting an increase in gratitude—namely, those that are *rationally accommodated* by the *hopes* present in the benefactor's initial manifestation of benevolence. When one acts from good will, one's benevolent desires may find expression not only in intentions and intentional action but in attitudes governed by weaker evidential norms—namely, hopes. The fittingness of gratitude growth is explained by the relation between the benefactor's benevolent hopes and the (unintended *but hoped-for*) state of affairs.

I proceed as follows. In section 1, I outline three paradigmatic components of gratitude. In section 2, I briefly discuss what it is to be more (or less) grateful. In section 3, I focus on a candidate case of gratitude growth. After identifying the mechanism by which gratitude growth occurs, I argue via contrastive analysis that gratitude growth is fitting for that subset of fortunate outcomes that are rationally accommodated by the manifestations of good will from which they proceed. Section 4 clarifies my proposal by comparing it to another possible view on which gratitude can be fittingly outcome sensitive and by addressing a concern about moral luck. The conclusion briefly reflects upon the implications that gratitude growth may have for questions concerning the rationality of emotional change.

1. GRATITUDE'S PARADIGMATIC FEATURES

To make sense of the change in gratitude described by gratitude growth, we need to get clear on gratitude's paradigmatic features. In this section, I propose that gratitude characteristically (1) represents a benefit to have been done *from good will*, (2) is a way of *appreciating* being benevolently benefited, and (3) *motivates* the beneficiary to give thanks.

I hasten to add that I am not committed to each of the abovementioned features being *essential* to gratitude such that, necessarily, when one is grateful, each feature is present. My aim here is not to provide an analysis of gratitude but rather to introduce those features of gratitude that do explanatory work in cases where gratitude fittingly increases, even if some are inessential to gratitude. I do, however, take the three abovementioned features to be present in paradigmatic cases of gratitude, understood as cases in which the benevolent agent is *successful* in her benevolent aim of benefiting another in a manner appreciated by the beneficiary.[13] The cases of gratitude that I discuss are paradigmatic in this sense.

13 Manela, "Does Gratitude to *R* for ϕ-ing Imply Gratitude that *R* ϕ-ed?" 3262n25. For skepticism concerning the mainstream view that gratitude is paradigmatically a response to benevolently motivated benefits (or attempted benefits), see Riedener, "Beyond Benefits."

1.1. Benevolence

In paradigm cases of gratitude, a feature of *S*'s being grateful to *V* for ϕ-ing is *benevolence*—i.e., *S* takes *V* to have ϕ-d from good will.[14] Gratitude construes another to have manifested good will or benevolence. To act benevolently is to act with the motivation of promoting another's good (whatever this is), for noninstrumental reasons. Good will can be unpacked in a variety of ways; we need not settle on a particular understanding. We might identify good quality of will with the "motive of duty" or with the intrinsic desire for what is good or right, properly conceptualized, or in some further way still.[15] Since I use the terms 'good will' (and 'good quality of will') and 'benevolence' interchangeably, benevolence too should be understood in this noncommittal sense.

I focus below on the subset of manifestations of good will featuring benevolent action. Benevolence might on some occasions be manifested exclusively in one's attentional and emotional responses (e.g., feeling distress upon recognizing another to be suffering). But in the paradigmatic case, willing good consists in part in benevolent action. For when it is within one's power to benefit another and when one's motives are also sufficiently strong, barring weakness of will and other agential failures, willing another's good is a matter of *exercising one's will to bring about* the other's good, i.e., to benefit her. Here, benevolence is efficacious, allowing for beneficence. In any case, I restrict myself below to manifestations of benevolence that, though they may include further elements, centrally feature beneficial action.

1.2. Appreciation

A paradigmatic feature of *S*'s being grateful to *V* for ϕ-ing is *appreciation*—i.e., that *S* be disposed to appreciate *V*'s ϕ-ing. Gratitude is a way of positively regarding that which it is about; it involves a positive evaluation of its object. As such, gratitude involves a construal of its object as *good* in some way. More particularly, gratitude paradigmatically involves appreciating another's benevolently given benefit. By 'appreciate,' I do not mean 'take pleasure in' or imply anything about hedonic tone. Gratitude *is* often understood to be a pleasant emotion.[16] But it is also sometimes characterized by (or at least associated

14 Where *S* and *V* stand for distinct agents and ϕ stands for some act token, as in "Sue (*S*) takes Val's (*V*'s) restarting Sue's stalled car (ϕ) to have been done from benevolence towards her."

15 Arpaly and Schroeder, *In Praise of Desire*.

16 See McCullough et al., "Is Gratitude a Moral Affect?"; Roberts, "The Blessings of Gratitude"; Tsang, "Gratitude and Prosocial Behaviour"; Watkins, *Gratitude and the Good Life*; Fagley, "The Construct of Appreciation"; and Darwall, "Gratitude as a Second-Personal Attitude (of the Heart)."

with) unpleasant feelings, e.g., those of indebtedness or guilt.[17] It may be that these negative feelings are not part of gratitude itself. Perhaps gratitude involves a disposition to be pleased, but this disposition is masked in certain circumstances. Though sympathetic to this idea, I take no official stand on it here.

In saying that gratitude involves a disposition to appreciate, I mean chiefly that gratitude is a way of *valuing* something, particularly another's manifestation of benevolence. I leave open the precise details of what valuing *X* consists in, although it includes being disposed to judge (or otherwise construe) *X* to be valuable.[18] Those (like me) sympathetic to the idea that valuing also involves a positive emotional disposition (which can be masked) can include this in their understanding of 'appreciation', but nothing hangs on this for my purposes.

Appreciation accounts for the intuition that a beneficiary who perceives another to have manifested good will but who remains indifferent (or worse) about this does not count as grateful. Notably, one may be grateful even if there are elements of another's manifestation of good will that one does not value. I am grateful, for example, to my young niece for making breakfast for me (or trying to), even if the product of her culinary efforts is inedible. As the saying goes, sometimes "it's the thought that counts." But that is to say, although I do not value the dish as such, I do *appreciate* (and so, construe as good) my niece's trying to do something kind for me.[19] Someone who appreciates *neither* the intended benefit nor "the thought" (i.e., another's benevolently aiming to benefit them) may say "thank you" (and otherwise act in ways the grateful person might), but I am reluctant to attribute the emotion of gratitude to them. After all, a speech act regularly used to express gratitude is "I appreciate it," and a sincerity condition of this kind of expressive speech act is that its speaker possesses the attitude purportedly expressed: that of appreciation.[20] Nevertheless, what matters for present purposes is that paradigmatic instances of gratitude involve appreciation.

17 See Gulliford et al., "Recent Work on the Concept of Gratitude in Philosophy and Psychology"; Manela, "Does Gratitude to *R* for ϕ-ing Imply Gratitude that *R* ϕ-ed?"; and Oh and Tong, "Mixed Emotional Variants of Gratitude."

18 Following Scheffler, valuing is standardly understood to consist partly in judging valuable ("Valuing").

19 If I *greatly disvalue* my niece's making of the dish—e.g., if it gives me severe food poisoning or if my niece destroys my oven in the process—this may preclude my being appreciative of her manifestation of good will. Though I am inclined to deny that I might be grateful nonetheless, I again take no stand here on whether gratitude requires appreciation (only that paradigmatic instances of gratitude involve appreciation).

20 The idea that gratitude's illocutionary aim includes expression of appreciation is prominent among speech act theorists, e.g., Searle, *Speech Acts*; and Leech, *Principles of Pragmatics*. Manela notably rejects that gratitude implies appreciation, but he means something

1.3. Motivation

A paradigmatic feature of *S*'s being grateful to *V* for ϕ-ing is *motivation*—i.e., that *S* is motivated to overtly give thanks to *V* for ϕ-ing. A grateful person not only appreciates another's benevolently benefiting her but, in appreciating the good will shown her, is motivated to express her gratitude *to* the benefactor.[21] The motivation is to overtly give thanks as, paradigmatically, the beneficiary not only wishes her benefactor well but wants to convey to the benefactor what the manifestation of benevolence means to her.[22] Sometimes all that is involved in a beneficiary's giving thanks is saying "thank you" (or some such) out of appreciative recognition of another's expressed good will. I am grateful, for example, to the person at the grocery store who invites me to check out ahead of him (his shopping cart overflowing with produce, mine containing a solitary case of coconut water). My gratitude is expressed with a smile and "thank you" in response to his offer—the smile and the utterance both being expressions of gratitude.[23]

But manifestations of good will sometimes have deeper effects on our motivational dispositions. The gratitude one feels toward one's mentor, parent, or rescuer may be considerably stronger and therefore apt to motivate one to do much more than say "thank you." In such cases, the motivation to give thanks might be construed as (something like) a *debt of gratitude*.[24] Leaving aside what so-called debts of gratitude amount to—including whether they should be understood on the model of duties—we can acknowledge that gratitude's motivational disposition may be experienced by the beneficiary as her owing a debt of gratitude to the benefactor.

Finally, giving thanks is not simply something a grateful agent is disposed to do but something she, at least from the perspective of her emotion, *has reason* to do.[25] When it is (at least *pro tanto*) justified, giving thanks is justified

---

narrower than I do by 'appreciation'—namely, that the beneficiary values the (intended) "benefit in and of itself" ("Gratitude and Appreciation," 285).

21 Gordon et al., "To Have and to Hold"; McCullough et al., "Is Gratitude a Moral Affect?"; Roberts, *Emotions*, 294; and Watkins, *Gratitude and the Good Life.*

22 See Shoemaker, "Qualities of Will," 117; Macnamara, "The Emotion of Gratitude and Communal Relationships," 108; and Smith, *The Theory of Moral Sentiments*, II.iii.10.

23 Sometimes we are not in any position to express our gratitude to our benefactors, e.g., because they are deceased or unknown to us. In these cases, one's motivation might manifest in wishing one could give thanks or in giving thanks to some proxy for one's benefactor. See Helm, "Gratitude and Norms," 186.

24 Lyons, "The Odd Debt of Gratitude"; McConnell, *Gratitude*; and Roberts, "The Normative and Empirical in the Study of Gratitude," 887.

25 See, e.g., Berger, "Gratitude"; McConnell, *Gratitude*, ch. 2; and Roberts and Telech, "The Emotion-Virtue-Duty Triad of Gratitude," 4.

by what gratitude is about—namely, one's being benevolently benefited by another. Indeed, the claim that beneficiaries sometimes construe themselves as owing debts of gratitude presupposes the idea that beneficiaries recognize *reasons* to give thanks (or at least, "seeming reasons," in cases of unfitting gratitude). While one might have other (e.g., instrumental) reasons for (or against) giving thanks—reasons irrelevant to whether one's addressee is gratitudeworthy—*reasons of gratitude* are those to which one is responsive to in being motivated to give thanks from recognition of being benevolently benefited by another.

To take stock, I characterize gratitude as a response that (1) represents a benefit as done *from good will*, (2) is a way of *appreciating* being benevolently benefited, and (3) *motivates* the beneficiary to give thanks. We need not assume these to be necessary features of gratitude, only that they are paradigmatic of gratitude.

2. BEING MORE (OR LESS) GRATEFUL

Implicit in the above outline is the idea that gratitude comes in degrees, or that it is a scalar phenomenon. Two persons may both be grateful, though one is *more grateful* than the other. Since changes in degrees of gratitude are central in what follows, we should pause on what it is for someone to be more (or less) grateful.

When one person (A) is more grateful than another (B) for X, there will be two dimensions along which the difference between A and B can be understood: an appreciative dimension and a motivational dimension. Along the appreciative dimension, the difference resides in A's being disposed to appreciate X more strongly. What is it for A to appreciate X more strongly than B appreciates X? Recall that appreciating something is a way of valuing that thing. Accordingly, whatever else it involves, A's appreciating X more strongly (or to a greater degree) than B does involves A's regarding X as more valuable than B does. If appreciating also involves a positive affective element—which I find plausible but do not presuppose—A's appreciating X to a greater degree than does B may consist also in A's having a stronger positive emotion concerning X than does B.

For A's gratitude to be stronger than B's along its motivational dimension is for A to be motivated to give thanks to a greater degree than B. This is ambiguous between two interpretations. On the first interpretation, the claim is exclusively about motivational strength; holding fixed what it is for one to "give thanks," if A has a stronger motivational disposition than B to give thanks, this means that, other things being equal, greater obstacles (or stronger countervailing reasons) are necessary to prevent A from giving thanks. On the second interpretation, being motivated to give thanks to a greater degree is a matter of being motivated to, as it were, *give more.* For example, thanking someone by taking them out to dinner is intuitively a "larger" (because, e.g., more resource-intensive) way of

giving thanks than, say, writing them a thank-you email. Which interpretation do I have in mind? Both. At least, a greater disposition to give thanks often involves both elements: a stronger motivational tendency to give thanks and a tendency to "give more."

While there is a sense in which gratitude increases if there is an increase in at least one of its appreciative or motivational elements, I understand gratitude growth conjunctively such that being "more grateful" implies an increase in *both* these dimensions.[26] There is at least something odd about a beneficiary's becoming more appreciative of her benefactor's manifestation of benevolence but not disposed to give thanks to any greater degree. (Directives to the effect that one "be more grateful" presumably call for greater thanksgiving, not merely greater appreciation.)[27] And there is something mysterious about one's gratitude increasing *only* along its motivational dimension. (What would account for this change if not an increase in one's appreciation?) I do not mean to deny the possibility of these kinds of (isolated) changes, but a view on which gratitude increases in both its appreciative and motivational elements is a stronger candidate for a view on which *gratitude* (and not merely one of its components) increases. Additionally, in Musical Fulfillment, Tim's gratitude *does* increase along both appreciative and motivational dimensions. (More on this shortly.) Accordingly, I seek to understand gratitude that grows in this way. On the descriptive story I provide in the following section, it is no accident that gratitude increases along both appreciative and motivational dimensions. The latter dimension is modulated by the former, at least in cases like Musical Fulfillment.

3. GRATITUDE GROWTH

In this section, I focus on the case presented in the introduction, Musical Fulfillment, which features Tim being grateful at one time (t_1) to his Aunt Ida and *more grateful* at a later time (t_2). This is a case in which the fortunate consequences of a manifestation of benevolence seemingly render fitting a greater degree of gratitude than was fitting in response to the initial manifestation of benevolence. For this proposal to be convincing, two aims must be met. First, we must specify the mechanism by which gratitude increases diachronically. *How is it* that Tim's gratitude grows? The second aim is to provide grounds for thinking that Tim's increase in gratitude is *fitting*. Since, to anticipate, not

26 Thanks to an associate editor for this journal and an anonymous referee for prompting me to think more carefully about this issue, which in turn contributed to how I understand gratitude feedback in section 3.1.

27 Whether such directives are *felicitous* (at least as naturally interpreted) is a further matter. See Macnamara, "'Screw You!' and 'Thank You'."

all fortunate consequences of manifestations of benevolence make gratitude growth fitting, meeting the second aim requires specifying which features of cases like Musical Fulfillment render gratitude growth fitting.

3.1. The Mechanism of Gratitude Growth: Gratitude Feedback

This subsection provides a description of the dynamics of Tim's change in gratitude. Recall that at t_1, Tim's response displays all paradigmatic features of gratitude$_1$: he takes his Aunt Ida to have benevolently benefited him in a way he appreciates and is motivated to give thanks in response. But given that Ida's manifestation of benevolence occurs at t_1, in virtue of what is Tim more grateful at t_2? Since there is no change in benevolence manifested between t_1 and t_2—and I stipulate that Tim knows this—Tim's gratitude growth cannot be understood as response to an increase in benevolence manifested. Instead, it is in virtue of his immense appreciation of having received the gift that his gratitude grows. Tim's appreciation that he received the flute greatly increases by t_2, and this appreciation *feeds back* into his already existing gratitude *to* his aunt *for* giving him the flute. Although Ida's benevolent action occurs prior to t_1, a richer appraisal of this action is available to Tim by t_2—e.g., "It is through Aunt Ida's benevolently gifting me a flute that I found my calling." While this description does not attribute *more benevolence* to Ida, it does attribute more value to Ida's benevolent action, via its outcome—namely, Tim's finding his calling. As it is gratitude's appreciative element that construes its object as valuable, we can understand why Tim's gratitude increases along its appreciative dimension by t_2. And as Ida, being Tim's benefactor, is the agent to whom Tim is disposed to give thanks, it is intelligible that Tim's motivation to give thanks to his benefactor increases with his increased appreciative disposition.[28]

I introduce the label *gratitude feedback* to refer to the kind of gratitude growth that Tim undergoes between t_1 and t_2. At t_2, Tim is motivated to express a greater degree of gratitude *to* his original benefactor for her gift, and this is plausibly so in virtue of an increase in Tim's gratitude along its appreciative dimension. Gratitude growth, as originally characterized, leaves open the mechanism by which gratitude increases; gratitude feedback specifies that the growth is a function of the increased appreciation of the benefit *feeding back* into the gratitude *to* the benefactor, generating a stronger motivation to give thanks.[29] While plenty questions concerning gratitude feedback remain, this rough outline suffices for now.

28 To put this in terms introduced by Zimmerman, Tim's gratitude increases in scope (it is *about more* at t_2 than it was at t_1) and in degree (he's *more grateful*, along gratitude's appreciative and motivational dimensions) ("Taking Luck Seriously").

29 Gratitude feedback, then, is *a type* of gratitude growth. There may be further types. Perhaps, for example, gratitude can increase in response to the *costs* that one's benefactor

In addition to its intuitive appeal for diagnosing the way in which Tim's gratitude increases, gratitude feedback has the following virtue: it appeals to a form of interaction among gratitude's elements that also accounts for counterfactual judgments in cases of *synchronic gratitude*. Intuitively, Beneficiary 1 might be more grateful to her benefactor (and so motivated to give thanks to a greater degree) than her counterfactual counterpart, Beneficiary 2, in virtue of the fact that Beneficiary 1 appreciates the manifestation of benevolence to a greater degree than does Beneficiary 2. Importantly, we can make sense of this difference in appreciation without assuming there is any difference in the degree of benevolence that each beneficiary attributes to her benefactor's action. Beneficiary 1 might be more appreciative if, for example, her values align more closely with her benefactor's (than Beneficiary 2's do with her psychologically identical benefactor) such that she is capable of recognizing the benefit she receives *as a benefit* (rather than merely an intended benefit) or (more simply) if her benefactor succeeds in benefiting her as intended, while her counterpart fails (through no shortcoming of benevolence). Leaving aside questions of fittingness, we typically *do* display more gratitude towards our benefactors for providing us with benefits that we greatly value than we do for benevolent *attempts* that are similarly motivated.[30] More generally, when comparing cases of synchronic gratitude across worlds, the degree to which one is disposed to give thanks can easily be understood as partly a function of the degree to which one appreciates the benefactor's manifestation of benevolence, holding fixed facts about perceived benevolence. While "the thought counts," it is not the *only* thing that counts, as far as gratitude is concerned. This so far is merely a descriptive claim. But if, in cross-world cases of synchronic gratitude, one's greater motivation to give thanks can be attributed to one's greater appreciation, it is natural to think that a single person's gratitude can change diachronically via the same basic mechanism—i.e., that in virtue of a difference (particularly an increase) in appreciation, one's motivation to give thanks increases.

I hasten to add that gratitude feedback is compatible with the following: after t_1 but long before t_2, Tim comes to better grasp the benevolence of his aunt's initial action and, resultantly, becomes more grateful to her for the original manifestation of good will. Tim is a child when he receives the gift, and he may be incapable of seeing the manifestation of good will for what it fully is. This may be because Tim has yet to occupy a role of guardianship over another.

comes to bear as a result of her beneficence. To develop this idea, see Manela, "Negative Feelings of Gratitude"; Miller, "Attributionism and Degrees of Praiseworthiness"; and Nelkin, "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness."

30 Gulliford and Morgan, "The Meaning and Valence of Gratitude in Positive Psychology," 205; Smith, *The Theory of Moral Sentiments*, II.iii.2.1; and Scanlon, *Moral Dimensions*, 151–52.

Once he is an adult (perhaps sooner), Tim may be capable of better grasping the value of Aunt Ida's manifestation of good will—i.e., value that was there all along but not visible, as it were, to his child self. This phenomenon—*enhanced benevolence perception*, as we might call it—is distinct from gratitude feedback. To illustrate, imagine Tim*, who is like Tim in all respects up to t_1 but who never takes flute lessons and does not become a world-renowned flautist. Now an adult, Tim* is better positioned to grasp Ida's benevolent motivations. Still, it is difficult to imagine him being as grateful to his aunt as is Tim, our musically fulfilled and world-renowned flautist. While Tim*'s maturity enables him to be grateful for features of the manifestation of benevolence that he could not apprehend as a child, his increased gratitude is responsive to the benevolence at t_1. There is no puzzle here. As an adult, he is simply in a better epistemic position to recognize the value of the initial manifestation of benevolence.

3.2. *A Fitting Change in Gratitude?*

Turning now to the question of fittingness, it is useful for dialectical purposes to introduce, as a foil of sorts, a skeptic about gratitude growth's fittingness whose skepticism has its basis in the idea that gratitude is fittingly sensitive *only to the quality of will manifested*. This skeptic—a quality of will theorist about gratitude—holds that while gratitude feedback may reflect something of *Tim's psychology* (and even of human psychology generally), it has no normative significance. "Perhaps people tend to be more grateful to their benefactors when their benevolent acts bear fruit," this skeptic might say, "but it hardly follows that they ought to be more grateful."

First, recall that my proposal is not that (1) Tim *should* be more grateful at t_2. More modestly, I maintain that (2) it is *fitting* for Tim to be more grateful at t_2. We can understand the sense of 'should' in 1 to be roughly that implying that Tim would be *criticizable for failing* to be more grateful at t_2. We can understand the sense of 'fitting' in 2 to be that which entails that Tim's gratitude is apt or correctly appraises its objects (that which it is about) but is not for that reason something one *ought* to feel. Proposal 2 is clearly stronger than 1. One can have a fitting reason to feel an emotion without it being the case that one would be criticizable for failing to have it. If Nour completes a charity marathon, she might be fittingly proud of herself for doing so. But it does not follow that she is criticizable if she fails to feel proud. To use concepts better suited to action, pride is in this case permissible but not required.[31] While I am sympathetic to the idea that Tim would be criticizably ungrateful if he did not undergo gratitude feedback, my goal here is only to make sense of it being *fitting* for Tim to be more grateful at t_2.

31 But see Fritz, "Why Fittingness Is Only Sometimes Demand-Like."

Granting the above point, the skeptic might contend that although Tim's increased gratitude at t_2 is unfitting, it may be otherwise appropriate (even all things considered), as Tim may have strong reasons other than those of fittingness to be more grateful at t_2. If there are, for instance, pragmatic reasons for Tim to be more grateful at t_2—reasons to be more grateful that have their basis in the benefits secured (or harms averted) by an increase in gratitude—it may be the presence of *these* reasons, not anything to do with the fittingness of increased gratitude, that makes it (and explains our intuition that it is) appropriate for Tim to be more grateful at t_2.

If given the opportunity, the skeptic might point to pragmatic considerations that could plausibly justify Tim's increase in gratitude at t_2. But rather than consider what is inevitably an incomplete set of possible pragmatic reasons for increased gratitude at t_2, by engaging in a kind of contrastive analysis, we can cast doubt more directly on the idea that the intuitive appropriateness of Tim's gratitude growth can be adequately accounted for without reference to fittingness. For we can identify cases in which gratitude feedback appears patently unfitting. But these cases differ markedly from the one at hand. Consider, for example, the following case.

> *Wealth and Happiness*: Tan's life until t_1 is in all relevant respects identical to Tim's. But suppose that around the time when Tim finds and starts playing with the flute that was given to him by his Aunt Ida, Tan instead trades his flute, which was given to him by his Aunt *Ada*, for a watch at a pawnshop. Sometime after, Julia stops Tan to compliment his watch. Julia and Tan hit it off. There begins a long and beautiful relationship. Early in this relationship, Tan and Julia develop the habit of playing the lottery, using a combination of their "lucky numbers." Ten years later, Tan and Julia renew their wedding vows, and the next day, they win millions in the lottery. At t_2, Tan recalls the role of the flute and is immensely grateful to Aunt Ada—much more so than he was at t_1. Thanks to her gift, he met the love of his life and, as a result, is also a millionaire. In addition to having the financial freedom to pursue a variety of exciting projects, Tan is part of a deeply gratifying relationship that provides him with a sense of happiness that he never thought possible as a young man.

For each of Tim and Tan, the flute occupies a causal role in the coming to be of their respective happy states at t_2. Let us stipulate that at t_2, Tim and Tan are equally grateful to their respective aunts, where this is a function of their being equally appreciative of the aspects of their adult lives that I have highlighted and their having equally strong motivations to give thanks. Yet there is intuitively an important difference between Musical Fulfillment, on the one hand, and Wealth

and Happiness, on the other. Were Tan to express towards his aunt gratitude of a magnitude like Tim's at t_2, I suspect that both we and his aunt would find something amiss in his response. It is of course understandable for the aunt to be pleased to learn of the unlikely effect of her gift. And Tan might of course be fittingly *glad* that he received the flute (although we would expect him to be especially glad that he traded it for the watch), but this attitude is distinct from the essentially agent-directed attitude that is gratitude. While gratitude feedback here might not be *unintelligible*, Tim's gratitude feedback is a much stronger candidate for fitting gratitude than is Tan's. A sign of this is found in the difficulty of imagining what form Tan's expression of gratitude at t_2 could take. To be sure, we can imagine Tan sharing some of his newfound wealth with his family, perhaps especially with his aunt. But, I submit, it will come more naturally to conceive of such actions as expressions of *generosity* rather than ways of expressing to his aunt what her initial manifestation of benevolence means to him—i.e., it is less natural to conceive of them as expressions of gratitude. Here, I am primarily reporting an intuition. Tim's gratitude feedback is intuitively fitting in a way that Tan's is not. I propose a normative basis for this intuition below.

3.3. *Hope and Rational Accommodation*

Taking at face value the intuition that Tim's gratitude feedback is fitting while Tan's is not, why might this be so? My proposal is that in Musical Fulfillment but not in Wealth and Happiness, the benevolence manifested at t_1 *rationally accommodates* the fortunate state of affairs at t_2. Although Aunt Ida does not intend or foresee that Tim will become a world-renowned flautist, we can nevertheless understand her to be *aiming at* Tim's musical success. Consider the conative attitudes underlying her gifting Tim the flute. What is it that she wants for Tim? Among other things, she wants Tim, for his own sake, to find artistic value and fulfillment in playing music. That is, she has an intrinsic desire that, roughly, Tim flourish and succeed musically.[32] While this is not itself something she can rationally *intend*—at least where intending that p implies believing or expecting that p (more on this below)—she *can* rationally intend to give Tim a flute and that Tim try his hand at the flute. Her benevolent desire for Tim's musical fulfillment is in this way actionable.[33]

32 Where to have an intrinsic desire for X is to desire X *at least in part* for its own sake.

33 More fully, Ida lacks sufficient evidence to believe that Tim will find artistic value and fulfillment in playing music, and intending that p (at least, *rationally* intending that p) implies believing/expecting that p. See Harman, "Practical Reasoning"; Davis, "A Causal Theory of Intending"; Setiya, *Reasons Without Rationalism*, 52; Wedgwood, "Instrumental Rationality," 294; and Velleman, "Practical Reflection" and "What Good Is a Will?" While I assume the standard view that intention entails belief/expectation, the claim that

But desires can find expression in conative attitudes other than intentions, e.g., in hopes. Given that Ida wants Tim to find artistic fulfillment in playing music and that she takes this desirable state of affairs to be *possible*, we can assume that this is something she hopes for. While it suffices for my purposes to stipulate that Ida has this hope, we can do better than this. Given the previously articulated features of Musical Fulfillment—along with facts about the nature of hope—attributing to Ida a hope for Tim's musical flourishing is not only intelligible but plausible. This is so for three reasons.[34] First, it is already assumed that Ida desires that Tim flourish and succeed musically. This desire underlies her benevolent intentional action of gifting Tim the flute. But given this desire, the main ingredient needed to attribute hope to Ida is the belief that it is *possible* that Tim flourish and succeed musically. For hope is widely thought to be constituted in part by the desire for some state of affairs, along with the belief that that state of affairs is possible (though not guaranteed).[35]

Second, just as desires, intentions, expectations, and beliefs can be implicit, so too with hopes.[36] Ida need not reflect on or attribute to herself a hope for Tim's musical flourishing. It is not difficult to imagine, however, that *were she asked* (around the time when she gifts Tim the flute—i.e., t_1) whether she hopes that Tim will thrive musically and find considerable value in playing the flute, she would reply affirmatively. Hope is a dispositional attitude, and this kind of (counterfactual) affirmative response, along with Ida's feelings of gratification in learning of Tim's success at t_2, are among the possible manifestations of Ida's hope.

Third, it is no part of my proposal that the content of Ida's hope is nearly as specific as what actually happens by t_2. Despite its lack of specificity, my hope (or desire) for, for example, a delicious cup of green tea is satisfied by my receiving an expertly brewed Saemidori cultivar gyokuro tea served in a ceramic cup. Indeed, the latter is a particular lofty way of satisfying my tea-related hope. Similarly, Tim's above-described t_2 state of affairs constitutes a specific way—a

---

it would not have been rational for Ida to intend Tim's musical success does not depend on this.

34 Thanks to an anonymous reviewer for prompting me to specify these reasons.

35 While hope has been analyzed as consisting in a combination of desire and belief (Hobbes, *Leviathan*, 36, I.VI.14; Day, "Hope"; Downie, "Hope"; and Milona, "Finding Hope"), many theorists think this "standard view" requires supplementation. See, e.g., Bovens, "The Value of Hope"; Pettit, "Hope and Its Place in Mind"; McGeer, "The Art of Good Hope"; Meirav, "The Nature of Hope"; Martin, *How We Hope*; and Chignell, "The Focus Theory of Hope." For an overview, see Rioux, "Hope."

36 Heckhausen, *Hoffnung und Furcht in der Leistungsmotivation*; Schultheiss and Köllner, "Implicit Motives."

particularly lofty way—of satisfying Ida's (far less specific) hope that Tim flourish and succeed musically.

Importantly, it can be rational for Ida to hope that Tim flourish musically even if it would not be rational for her to have an intention with the same content. For assuming that intending that *p* implies believing or expecting that *p* will obtain—such that one assigns a greater probability to *p* than not-*p* (i.e., one assigns a subjective probability of greater than 50 percent to *p*)—*hoping* that *p* implies believing only that the desired state of affairs *is possible*, even if the subjective probability one assigns to it is quite low, e.g., 1 percent. So although she does not (and rationally cannot) intend Tim's musical fulfillment at t_1, a hope of this sort is rationally available to her at t_1. Further, given that her intrinsic desire for Tim's musical fulfillment is sufficiently strong to find expression in her benevolent intentions and actions (those that render fitting gratitude at t_1), it is not only possible but plausible that her intrinsic desire would also give rise to the hope in question.[37]

Next, while hope is not action guiding in the same way as intention, it can nonetheless be expressive of one's quality of will.[38] Without settling on a view of what is essential to hope, note that in contrast to merely desiring some state of affairs, hoping is often thought to be broadly agency-involving—e.g., in including endorsement of one's desire, the disposition to devote mental energy imagining what the world would be like if one's hope were realized, etc. That we think of hope as expressive of an agent's good will is reflected too in our responses to others' hopes concerning our well-being. When others wish us well, they are often expressing interpersonal hopes, as in, for example, "I wish you a speedy recovery" or "I hope you get well soon." These "well wishes" are often met with thanks. While such exchanges may sometimes be mere pleasantries, they can also be heartfelt, both by the giver and by the receiver.

My proposal, then, is that Ida's benevolent attitudes at t_1 rationally accommodate the happy state of affairs at t_2. In particular, it is the fulfillment of her hope for Tim's musical success that accounts for the fitting *increase* of Tim's gratitude. While her benevolent intentions are fulfilled long before, her benevolent hope becomes satisfied only by (or around) t_2. The benevolent intentions (to give Tim the flute and that he try his hand at the flute) and her associated

37 In referring to desires as "finding expression in" intentions, I aim to remain neutral on the metaphysics of intention—e.g., on whether intentions are partly constituted by desires or are instead *sui generis* attitudes.

38 On quality of will being expressed through nonactions and nonintentions, in a way that licenses reactive emotions, see Shoemaker, *Responsibility from the Margins*; Smith, "Responsibility for Attitudes"; and Zagzebski, *Exemplarist Moral Theory*, 231. On hope as expressive of good will among friends, see also Scanlon, *Moral Dimensions*, 132.

actions rationally accommodate gratitude to the degree fitting at t_1, and the increase in fitting gratitude is accounted for by the hope's fulfillment at a much later time. It is in this sense that the benevolent attitudes at t_1 rationally accommodate the happy state of affairs at t_2. The total degree of gratitude fitting at t_2, then, is accounted for by the suite of satisfied benevolent attitudes underlying her gifting Tim the flute at t_1—namely, the intentions and hope mentioned above. These attitudes are of a piece in that they share a conative basis in the intrinsic desire aimed at Tim's musical fulfillment.

Before specifying what must hold of some hope and some state of affairs such that the former rationally accommodates the latter, let me emphasize that it is not the mere benevolent hope that renders fitting the increase in gratitude but rather that hope's being *fulfilled* in the right kind of (to-be-further-specified-below) way. After all, this hope is present at t_1. And although the presence of the benevolent hope at t_1 is a key ingredient for the fittingness of a greater degree of gratitude at t_2, it is insufficient. This can be illustrated by noting that Tan's Aunt Ada has the same hope (i.e., a type-identical hope).[39] But Tim's gratitude feedback is intuitively fitting in a way that Tan's is not. The relevant difference, on my view, is that Ida's hope is fulfilled—i.e. as I put it, her benevolent hope is not only manifested but *realized*, and it is not until t_2 that it is realized.

To help motivate the idea that the realization of a benevolent attitude can impact gratitude's fittingness, recall that it was not Ida's mere benevolent intentions that rendered gratitude fitting (to the degree that it was) at t_1. Her intentional *actions* and intended *outcomes*—that is, the "realizations" of her benevolent intentions—form an integral part of the object of Tim's gratitude at t_1. And indeed, quality of will theorists generally include such worldly realizations of benevolent attitudes among "manifestations of quality of will."[40] As Strawson writes, reactive attitudes are "reactions to the quality of others' wills towards us, as *manifested in their behavior*."[41] We do not always succeed in manifesting our quality of will as intended, so some such manifestations of good will may be better characterized as *attempts*. Nevertheless, the quality of will that others have toward us is of interpersonal significance in large part owing

39 Below, I consider the idea that Ida's and Ada's hopes might be nonidentical owing to their outcomes.

40 This is true of not only Strawson (in "Freedom and Resentment") and subsequent Strawsonians (e.g., McKenna, *Conversation and Responsibility*, 59; Macnamara, "Holding Others Responsible," 89; Shoemaker, "Qualities of Will" and *Responsibility from the Margins*; and Rosen, "The Alethic Conception of Moral Responsibility," 77) but other quality of will theorists as well (e.g., Arpaly, *Unprincipled Virtue*, 84; and Arpaly and Schroeder, *In Praise of Desire*, 170).

41 Strawson, "Freedom and Resentment," 84.

to its material role in our various relationships. A quality of will theorist, then—according to whom gratitude is fittingly sensitive only to *manifested* quality of will—*already* includes benevolently motivated outcomes among the manifestations of good will that are gratitude's proper objects. For this reason, though introduced as a foil, the quality of will theorist *can*, provided she adopts the idea that hopes can be expressive of quality of will, accommodate my proposal that gratitude feedback can be fitting. For as articulated above, benevolent hopes can be understood as part of the benefactor's manifestation of good will.

Since in Musical Fulfillment, the relevant benevolent hope is present at t_1 yet operative in a different way at t_2, it is useful to distinguish between (1) quality of will that is (merely) manifested and (2) quality of will that is *realized*, where realization implies the fulfillment of a motivational attitude (previously) manifested. Thus, we can say that in our lead example, although the *manifestation* of quality of will is fixed in the past (i.e., at some time prior to t_1), an element of the quality of will manifested at t_1—namely, a benevolent hope—becomes realized owing to the fortunate state of affairs at t_2. In saying that the hope-involving element of Ida's t_1 manifestation of benevolence is *realized* at t_2, I am simply providing a way of describing, in terms of quality of will, the change that occurs when Ida's benevolent hope becomes satisfied such that it *rationally accommodates* the happy state of affairs of Tim's musical fulfillment.

One might wonder whether the present proposal can avoid the verdict that Tan's gratitude feedback is also fitting. After all, can we not suppose that Aunt Ada had hoped that, say, the flute would bring Tan joy? If so, given that the flute *did* bring him joy, albeit in an unintended way (in leading to his attainment of the watch that led to his wealthy and happily married life at t_2), am I not committed to saying that Tan's happy state of affairs at t_2 is rationally accommodated by the benevolent attitudes underlying the gift at t_1?

I am not committed to this. For intentions place rational constraints on one's hopes.[42] Recall that what is true of Aunt Ida at t_1 is true too of Aunt Ada at t_1. Accordingly, at t_1, Aunt Ada intends to gift Tan the flute—but *crucially*, this intention has its conative basis in the intrinsic desire that Tan flourish musically. But the hope that one's nephew be happy *in any way whatsoever* cannot be expressive of the intrinsic desire that one's nephew flourish musically. This can be illustrated by attending to the diverging satisfaction conditions of these conative attitudes. Indeed, Tan's happy state of affairs at t_2 involves the satisfaction of the former but not the latter desire. But even if we modify Wealth and Happiness such that Ada *does* have an unconstrained hope of the above sort

42 On the rational constraints that intentions impose on our beliefs and other intentions, see Bratman, *Intention, Plans, and Practical Reason*. On rational relations between hopes and intentions, see Mason, "Hoping and Intending," 526–27.

(that the flute make Tan happy, in any way whatsoever), assuming we keep fixed the other features of the case, this hope cannot rationally accommodate Tan's state of affairs at t_2. For a benefactor's benevolent hope to rationally accommodate some state of affairs, that hope must cohere with the benevolent aim underlying the benefactor's intention. Hope is, in this way, governed by a norm of structural rationality. Since the flute leads to Tan's happiness in a way that conflicts with the aim of his musical flourishing, this state of affairs cannot be rationally hoped for in the sense at issue. Thus, Tan's happy state of affairs at t_2 is not rationally accommodated by the manifestation of benevolence at t_1.[43] The above point allows us to clarify that for a hope to rationally accommodate some state of affairs, the hope must be rational not only in being epistemically warranted (such that one is justified in believing that the hope's satisfaction is *possible*) but in being coherent with one's intentions.

We are in a position now to provide a definition of rational accommodation for hope.

> *Rational Accommodation*: For a hope (H) to rationally accommodate some state of affairs (A), (1) H must be epistemically rational, (2) H must display structural rationality, and (3) A must constitute the fulfillment of H.

Next, for a hope to alter the fittingness of one's degree of gratitude, not only must it rationally accommodate some state of affairs; the hope must be benevolent. (Gratitude, after all, is fittingly responsive not to quality of will generally

43 We *can* modify Wealth and Happiness such that Tan's gratitude feedback is fitting. Suppose that the desire underlying Ada's manifestation of benevolence at t_1 was *not* that Tan flourish musically but that Tan *be happy*. This would be a substantive modification (as it would change the benefactor's intrinsic desire), and assuming that at t_1, the cases of Tim and Tan are identical in all normatively relevant respects, in this modified case, *Tim's* fortunate state of affairs at t_2 would not be rationally accommodated by his benefactor's hope, except perhaps to the extent that his flourishing musically is a way of (or also involves his) being happy. Still, given my aim of establishing that gratitude feedback can be fitting owing to the satisfaction of the benefactor's benevolent hopes, I can grant that Tan's gratitude feedback *can* be fitting in a substantively modified version of Wealth and Happiness. If the aunt's benevolent desire is quite generic (and more easily satisfiable), however—as is the desire that one's nephew be happy *in any way whatsoever*—it may be that *at* t_1, she is considerably less gratitudeworthy for gifting Tan the flute than Ida is in Musical Fulfillment at t_1 (i.e., where Ida's desire is that her nephew flourish musically). For it may be that the satisfaction conditions of the benevolent attitude provide an upper limit to the degree of benevolence one can manifest in expressing it. Cf. Arpaly and Schroeder, *In Praise of Desire*, 189. If so, it may be that the degree to which Tan's gratitude (in the modified case) may fittingly grow is more limited than the degree to which Tim's gratitude may fittingly grow.

but to *good* quality of will—i.e., benevolence.)[44] Additionally, my proposal is only that a rationally accommodating benevolent hope can augment the degree of fitting gratitude *in cases where* this hope is expressive of the same intrinsic desire underlying the benevolent intentions from which the benefactor originally acted. In such cases, the benevolent intentions and hopes are of a piece; they are constituents of a single manifestation of good will. Ida's manifestation of good will includes both (1) her benevolent intentions (manifested and realized by t_1) and (2) her benevolent hopes (manifested at/before t_1 but realized only at/around t_2). Underlying both 1 and 2 is the intrinsic desire that Tim flourish and succeed musically.

The present proposal has the resources to explain why Tim's gratitude growth is intuitively fitting, while Tan's is not. Even if Tan somehow acquires equal or stronger overall reason to be more grateful to his aunt, the cases of Musical Fulfillment and Wealth and Happiness remain importantly different. Suppose, for example, that Tan has strong pragmatic reasons to be more grateful to his aunt at t_2. Perhaps doing so will help launch Ada out of a debilitating depression. The positing of such reasons only brings into starker contrast the difference between intuitively fitting and unfitting cases of outcome-sensitive gratitude. Tim's gratitude feedback is intuitively fitting in the absence of any pragmatic considerations. He might foresee that his aunt will be gratified by his t_2 expression of gratitude, but this is not why his gratitude increases at t_2. His gratitude remains importantly anchored in backward-looking considerations; it is gratitude to Ida for the benevolently given gift, but now the gift is construed as that which has led to Tim's musical success and flourishing. This change—particularly the *increase* in Tim's gratitude—is fitting given that the fortunate outcome of the gift is rationally accommodated by the benevolent attitudes (particularly, the benevolent hope) manifested at t_1. Tan's happy state of affairs at t_2 is, by contrast, not rationally accommodated by the benevolent attitudes manifested at t_1. The relation between the manifestation of benevolence at t_1 and Tan's happy state of affairs at t_2 is merely causal. The skeptic about fitting gratitude feedback, however, lacks such grounds for distinguishing between Tim and Tan. Though the skeptic can say that the aunts are fitting targets of gratitude at t_1 for their benevolently given gifts, the fruitfulness of a manifestation of benevolence is the wrong sort of thing to alter the fittingness of gratitude, and so Tim's gratitude feedback is, according to the skeptic, just as unfitting as is Tan's. This verdict, however, is neither intuitively plausible nor required for those attracted to the view that gratitude is fittingly responsive

44 As stated, Rational Accommodation is compatible with *malevolent* (or morally neutral) hopes accommodating some state of affairs. I briefly return to the possible import of this in section 5 below.

to manifestations of good quality of will. For benevolent hopes are plausibly understood as manifestations of good quality of will.

Before considering an alternative way that gratitude might be understood as fittingly outcome sensitive, let me address a concern about moral luck, particularly resultant moral luck (or "luck in consequences").[45] One might worry that gratitude feedback entails resultant moral luck and that it should be resisted on those grounds. Ida and Ada manifest the same degree of benevolence (and exercise the same degree of volitional control) in gifting their respective nephews a flute. But if from t_2 onward, Ida is the fitting target of more gratitude than is Ada, given that this is explained by factors external to quality of will, it appears that Ida's being more gratitudeworthy is a matter of luck. One way to press this point is to claim that it is *unfair* to Ada that she be the fitting target of less gratitude than Ida, given that each manifests the same degree of good will (and exercises the same degree of control) in benefiting their respective nephews.

In response, the moral luck debate traditionally concerns *moral responsibility*—that is, blameworthiness and praiseworthiness. Even if responding with gratitude can be a way of praising, it is far from uncontroversial—and in my view false—that being (more) gratitudeworthy entails being (more) praiseworthy.[46] In any case, I make no such assumption here. One reason to doubt this assumption is that gratitudeworthiness is intuitively patient-relative in a way that praiseworthiness is not. At t_2, Ida is the fitting target of more gratitude *from Tim*, given that Tim is the beneficiary of Ida's benevolent action. But this does not entail that it would be fitting for third parties to respond to Ida with a similarly heightened degree of a corresponding positive attitude, like admiration or some other impersonal praising response.[47] For all I have said, Ada and Ida might be *equally* (and unchangingly) praiseworthy. Thus, gratitude feedback does not entail resultant moral luck.[48]

45 Nagel, "Moral Luck," 27.

46 However, see Scanlon, *Moral Dimensions,* 151–52, for the view that (1) blame's contrary is gratitude (rather than praise) and a qualified defense of the idea that (2) successful manifestations of good will render fitting more gratitude (or at least more by way of *grateful acknowledgement*—which I take to be a way of giving thanks—if not a greater readiness to help one's benefactor) than their unsuccessful counterparts. This is introduced as a point not about gratitude growing but of cross-world differences in synchronic gratitude. See section 3.1 above.

47 But on third-party gratitude that is proposed to be independent of one's identifying with the beneficiary and is meant to be distinct from both praise and admiration, see Eskens, "Moral Gratitude."

48 There are, however, more capacious views of moral luck, according to which, moral luck exists if factors beyond one's control (or external to one's quality of will) can directly determine one's positive (or negative) *moral status*, where one's positive moral status may include

4. RETROSPECTIVE DETERMINATION

Before I conclude, let us consider another way in which gratitudeworthiness might be fittingly outcome sensitive. Drawing on work by Gerald Lang, one might maintain that until a benevolent outcome materializes, there is no determinate level of gratitude that is fitting.[49] For the moral goodness of the benefactor's benevolent attitudes may remain unfixed until its outcomes materialize. While I characterize Ida and Ada as acting on identical attitudes in benefiting their respective nephews, it may be that the *goodness* of the attitudes "embodied in these agents' acts is fixed retrospectively," and as such, the goodness of their respective motivations is unequal.[50] Gratitudeworthiness, then, might be outcome sensitive owing to the truth of the "determination claim," according to which the goodness of an actual outcome caused by moral agents who act on morally good mental states can fix the degree of goodness of the mental states that were embodied in these agents' acts.[51]

There are two ways of construing the determination claim in the present context. Construal 1 is consistent with my account of gratitude growth. Construal 2 is incompatible with my account, but it is also implausible. On Construal 1, the outcomes of Ida's good mental states at t_1 play a role in fixing the goodness of her mental states and also the degree to which she is gratitudeworthy at t_1. While I have reserved talk of outcomes for the unforeseen and unintended results of manifestations of benevolence, if we understand outcomes to include results that the benefactor intends, outcomes are on the scene at t_1. We can use *proximate outcomes* to refer to intended outcomes. Among the proximate outcomes of Ida's benevolence, suppose, is Tim's being pleased with the flute—that is, that receipt of the flute bring him some pleasure (though in a way that is constrained with his engaging with the flute *qua* musical instrument). Tim, as it happens, is very pleased to receive the flute. This, then, is an outcome Ida intends. But we can

properties beyond one's being (more) praiseworthy, e.g., being (more) gratitudeworthy, being forgiven. See Telech, "Forgiveness and Moral Luck," 239. Cf. Story, "Interpersonal Moral Luck and Normative Entanglement"; and Hartman, "Kant Does Not Deny Resultant Moral Luck," 144. Further discussion of the idea that gratitude feedback may imply the existence of positive resultant moral luck (capaciously understood) must wait for another time.

49 To be clear, Lang limits his discussion to blameworthiness (*Strokes of Luck*). Nevertheless, it is worth considering how his "determination claim" might work in the context of gratitude.

50 Lang, *Strokes of Luck*, 63.

51 My characterization of the determination claim is adapted from Lang, *Strokes of Luck*. He endorses this claim only in the context of blameworthiness and morally flawed mental states. The original sentence reads: "The badness of an actual outcome caused by moral agents who act on morally flawed internal states can help to fix the degree of badness of the mental states which were embodied in these agents' acts" (62).

imagine a nearby version of the case in which Ida has the same benevolent intention (and same motivating attitudes generally), yet the outcome is less successful.

Suppose our less fortunate benefactor—call her Aunt Uma—receives from the shop a flute made of a metal to which her nephew, Tum, is mildly allergic. Tum, we can suppose, is pleased to receive the flute—and so Aunt Uma's intended outcome *does* materialize—but Tum is *less pleased* to receive his flute than Tim is to receive his. (Suppose Tum knows that after about thirty minutes of handling the instrument, he will develop a mild and temporary rash.) Though I take no official stand on this issue, it may be that the goodness of Ida's and Uma's respective intentions is fixed in part by the outcomes to which they give rise. That is, it may be that Uma's benevolent intention is *less good* than is Ida's owing to its being embodied in a less desirable outcome. Assuming that one's gratitudeworthiness is in part a function of the goodness of one's motivating attitudes, Uma may be the fitting target of *less* gratitude than is Ida. This, however, is a claim about gratitudeworthiness *at* t_1. As such, it is wholly compatible with the claim that *at* t_2, Ida is the fitting target of a greater degree of gratitude than she was at t_1, and this is so owing to the outcome at t_2 being rationally accommodated by her t_1 hope. But as Musical Fulfillment and Wealth and Happiness are identical at t_1, Construal 1 of the determination claim fails to differentiate between the two cases.

On Construal 2 of the determination claim, Ida's level of gratitudeworthiness is indeterminate *until* t_2. I take lack of determinacy here to imply that until t_2, there is no fact of the matter concerning the level of Ida's gratitudeworthiness. After all, if 'indeterminate' just means mutable, then that is consistent with my view. On Construal 2, then, Ida's gratitudeworthiness does not *increase* between t_1 and t_2. Rather, it is only from t_2 onward that there is a fact of the matter concerning Ida's gratitudeworthiness. This construal of the determination claim is at the very least counterintuitive. By Tim's and Ida's lights, Tim's grateful response at t_1 *is* fitting. But the claim that until t_2 there is no fact of the matter concerning Ida's gratitudeworthiness implies that Tim's gratitude at t_1 is not fitting, as there are no fit-making facts at that time. Additionally, in worlds in which the aunt's hope is not fulfilled (e.g., the world of Wealth and Happiness), it does not seem plausible that we need to wait until t_2 for there to be a fact of the matter concerning the aunt's gratitudeworthiness. A more natural understanding of the case is that Aunt Ida's gratitudeworthiness at t_2 is greater than her degree of gratitudeworthiness a t_1—i.e., it increases. But an increase in gratitudeworthiness requires that there was a fact of the matter concerning Ida's gratitudeworthiness at t_1.

While Construal 2 of the determination claim is coherent, its implication that there is no fact of the matter concerning Ida's gratitude until t_2 (roughly

twenty years after the original manifestation of benevolence at t_1) renders its way of accounting for the outcome sensitivity of gratitudeworthiness revisionary in a way that the rational accommodation view is not. Nevertheless, if on Construal 2 of the determination claim, the goodness of Ida's t_1 hope is taken to be retroactively fixed by its fulfillment at t_2, then an important feature of my view—that the fulfillment of benevolent hopes can contribute to one's overall degree of gratitudeworthiness—is granted.

5. CONCLUSION

Gratitude, I have argued, is fittingly sensitive not only to the benevolence others manifest in their actions toward us but also to a subset of the beneficial consequences that such benevolent actions may bring about. Focusing on diachronic gratitude, I have presented a case of gratitude feedback: a phenomenon wherein the fortunate consequences of a manifestation of benevolence render intuitively fitting a greater degree of gratitude than did the initial manifestation of benevolence. For gratitude feedback to be fitting, I have argued, the initial manifestation of benevolence must play not only a causal role in the production of the fortunate outcome; that outcome must be rationally accommodated by the manifestation of good will, particularly by the satisfaction of a benevolent hope that shares a conative basis with the intentions of the manifestation of good will. For this kind of hope to rationally accommodate the state of affairs, in addition to its being fulfilled by that state of affairs, the hope must be epistemically rational (such that its component belief in the possibility of the hoped-for state of affairs is *justified*), and it must display structural rationality, such that it coheres with the intentions of the manifestation of good will.

In saying something about gratitude, I have said something about emotional change. Recall the general question concerning the rationality of changes in backward-looking emotions, changes that occur, and are intuitively fitting, despite the unchanging nature of that which seemed to render fitting the emotion in the first place. The literature on the fittingness of changes in backward-looking emotions tends to focus on the diminishment of emotion. In providing a vindicatory account of gratitude growth, we recharacterize the problem as encompassing questions both of emotional diminishment and of emotional growth.

Looking forward, there are questions to consider concerning the interplay between the fittingness of emotional growth and diminishment. It may be that a token backward-looking emotion can fittingly grow and then fittingly diminish (or vice versa). The idea of feeling increasingly sad about some loss that one eventually ceases to feel sad about, for instance, is far from unintelligible.

Additionally, while gratitude growth's fittingness does not straightforwardly entail the fittingness of a negative counterpart ("resentment growth"), if a case can be made for resentment growth—perhaps using the resources here developed, e.g., by appealing to the significance of the satisfaction of a *malevolent* hope underlying one's wrongdoing—then assuming that augmented responses of resentment may eventually diminish and subside (e.g., via a process of forgiveness), then there too we will have occasion to reflect on the interplay between the fittingness of growth and diminishment in a single emotion.[52]

University of the Fraser Valley
daniel.telech@ufv.ca

REFERENCES

Archer, Alfred, and Benjamin Matheson. "Admiration over Time." *Pacific Philosophical Quarterly* 101, no. 4 (2020): 669–89.

Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, 2003.

Arpaly, Nomy, and Timothy Schroeder. *In Praise of Desire*. Oxford University Press, 2014.

Berger, Fred. "Gratitude." *Ethics* 85 (1975): 298–309.

Bovens, Luc. "The Value of Hope." *Philosophy and Phenomenological Research* 59 (1999): 667–81.

Bratman, Michael. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

Callard, Agnes. "The Reason to Be Angry Forever." In *The Moral Psychology of Anger*, edited by Owen Flanagan and Myisha Cherry. Rowman and Littlefield, 2017.

Camenisch, Paul. "Gift and Gratitude in Ethics." *Journal of Religious Ethics* 9 (1981): 1–34.

Chignell, Andrew. "The Focus Theory of Hope." *Philosophical Quarterly* 73, no. 1

52 Earlier versions of this article benefited from discussion at the University of Bern, University of Chicago, and Hebrew University of Jerusalem. For helpful comments and questions, I would like to thank Agnès Baehni, Constant Bonnard, Agnes Callard, Andreas Brekke Carlsson, Rebecca Chan, Simon-Pierre Chevarie-Cossette, David Enoch, Hadi Fazeli, Vincent Grandjean, David Heyd, Leora Dahan Katz, Andrew Khoury, Brian Leiter, Ben Matheson, Oded Na'aman, Paul Russell, Marya Schechtman, Shlomi Segal, and Dave Shoemaker. Special thanks to Max Lewis and Olle Blomberg for very helpful written comments on more recent versions of the article. Finally, comments from associate editors, along with comments from several anonymous reviewers, were very useful in improving the article.

(2023): 44–63.
Cholbi, Michael. "Grief's Rationality, Backward and Forward." *Philosophy and Phenomenological Research* 94, no. 2 (2017): 255–72.
Coren, Daniel. "Giving Up Gratitude." *Analytic Philosophy* 66, no. 1 (2025): 22–36.
D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65–90.
———. *Rational Sentimentalism*. Oxford University Press, 2023.
Darwall, Stephen. "Gratitude as a Second-Personal Attitude (of the Heart)." In Roberts and Telech, *The Moral Psychology of Gratitude*.
Davis, Wayne. "A Causal Theory of Intending." *American Philosophical Quarterly* 21 (1984): 43–54.
Day, John. "Hope." *American Philosophical Quarterly* 6 (1969): 89–102.
Downie, Robin. "Hope." *Philosophy and Phenomenological Research* 24 (1963): 248–51.
Emmons, Robert, and Michael McCullough. "Counting Blessings Versus Burdens: An Empirical Investigation of Gratitude and Well-Being in Daily Life." *Journal of Personality and Social Psychology* 84 (2003): 377–89.
Eskens, Romy. "Moral Gratitude." *Journal of Applied Philosophy* 42, no. 1 (2025): 115–30.
Fagley, Nancy. "The Construct of Appreciation: It Is So Much More than Gratitude." In *Perspectives on Gratitude: An Interdisciplinary Approach*, edited by David Carr. Routledge, 2016.
Fritz, James. "Why Fittingness Is Only Sometimes Demand-Like." *Philosophical Studies* 180, no. 9 (2023): 2597–616.
Gordon, Amie M., Emily A. Impett, Aleksandr Kogan, Christopher Oveis, and Dacher Keltner. "To Have and to Hold: Gratitude Promotes Relationship Maintenance in Intimate Bonds." *Journal of Personality and Social Psychology* 103, no. 2 (2012): 257–74.
Gulliford, Liz, and Blaire Morgan. "The Meaning and Valence of Gratitude in Positive Psychology." In *The Routledge International Handbook of Critical Positive Psychology*, edited by Nicholas Brown, Tim Lomas, and Francisco Jose Eiroa-Orosa. Routledge, 2017.
Gulliford, Liz, Blaire Morgan, and Kristján Kristjánsson. "Recent Work on the Concept of Gratitude in Philosophy and Psychology." *Journal of Value Inquiry* 47, no. 3 (2013): 285–317.
Harman, Gilbert. "Practical Reasoning." *Review of Metaphysics* 29 (1976): 431–63.
Hartman, Robert. "Kant Does Not Deny Resultant Moral Luck." *Midwest Stud-*

ies in Philosophy 43, no. 1 (2019): 136–50.
Heckhausen, Heinz. *Hoffnung und Furcht in der Leistungsmotivation* [Hope and Fear Components of Achievement Motivation]. Anton Hain, 1963.
Helm, Bennett. "Gratitude and Norms: On the Social Function of Gratitude." In Roberts and Telech, *The Moral Psychology of Gratitude*.
Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55.
———. "The Wrong Kind of Reason." *Journal of Philosophy* 102, no. 9 (2005): 437–57.
Hobbes, Thomas. *Leviathan*. Edited by John Gaskin. Oxford University Press, 1998.
Lang, Gerald. *Strokes of Luck: A Study in Moral and Political Philosophy*. Oxford University Press, 2021.
Leech, Geoffrey. *Principles of Pragmatics*. Longman, 1983.
Lewis, Max. "Gratitude: Its Nature and Normativity." *Philosophy Compass* 19, no. 8 (2024): e13015.
Lyons, Daniel. "The Odd Debt of Gratitude." *Analysis* 29, no. 3 (1969): 92–97.
Macnamara, Coleen. "The Emotion of Gratitude and Communal Relationships." *Journal of Applied Philosophy* 42, no. 1 (2025): 96–114.
———. "Holding Others Responsible." *Philosophical Studies* 152, no. 1 (2011): 81–102.
———. "'Screw You!' and 'Thank You.'" *Philosophical Studies* 165, no. 3 (2013): 893–914.
Manela, Tony. "Does Gratitude to *R* for ϕ-ing Imply Gratitude that *R* ϕ-ed?" *Philosophical Studies* 177, no. 11 (2020): 3245–62.
———. "Gratitude and Appreciation." *American Philosophical Quarterly* 53, no. 3 (2016): 281–94.
———. "Negative Feelings of Gratitude." *Journal of Value Inquiry* 50, no. 1 (2016): 129–40.
Martin, Adrienne. *How We Hope: A Moral Psychology*. Princeton University Press, 2014.
Marušić, Berislav. "Do Reasons Expire? An Essay on Grief." *Philosophers' Imprint* 25 (2018): 1–21.
Mason, Cathy. "Hoping and Intending." *Journal of the American Philosophical Association* 7, no. 4 (2021): 514–29.
Mason, Sam. "Dimensions of Emotional Fit." *Philosophical Quarterly* 75, no. 1 (2025): 125–46.
McAleer, Sean. "Propositional Gratitude." *American Philosophical Quarterly* 49, no. 1 (2012): 55–66.
McConnell, Terrance. *Gratitude*. Temple University Press, 1993.

McCullough, M. E., S. D. Kilpatrick, R. A. Emmons, and D. B. Larson. "Is Gratitude a Moral Affect?" *Psychological Bulletin* 127, no. 2 (2001): 249–66.

McGeer, Victoria. "The Art of Good Hope." *Annals the American Academy of Political and Social Science* 592, no. 1 (2004): 100–27.

McKenna, Michael. *Conversation and Responsibility*. Oxford University Press, 2012.

Meirav, Ariel. "The Nature of Hope." *Ratio* 22 (2009): 216–33.

Miller, Daniel. "Attributionism and Degrees of Praiseworthiness." *Philosophical Studies* 179, no. 79 (2022): 3071–87.

Milona, Michael. "Finding Hope." *Canadian Journal of Philosophy* 49, no. 5 (2019): 710–29.

Moller, Dan. "Love and Death." *Journal of Philosophy* 104, no. 6 (2007): 301–16.

Na'aman, Oded. "The Fitting Resolution of Anger." *Philosophical Studies* 177, no. 8 (2020): 2417–30.

———. "The Rationality of Emotional Change: Toward a Process View." *Noûs* 55, no. 5 (2021): 245–69.

Nagel, Thomas. "Moral Luck." In *Mortal Questions*. Cambridge University Press, 1979.

Nelkin, Dana. "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness." *Noûs* 50, no. 2 (2016): 356–78.

Oh, Vincent, and Eddie Tong. "Mixed Emotional Variants of Gratitude." *Cognition and Emotion* 37, no. 3 (2023): 572–58.

Peterson, Christopher, and Martin Seligman. *Character Strengths and Virtues: A Handbook and Classification*. Oxford University Press, 2004.

Pettit, Philip. "Hope and Its Place in Mind." *Annals of the American Academy of Political and Social Science* 592 (2004): 152–65.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-Attitude and Value." *Ethics* 114, no. 3 (2004): 391–423.

Riedener, Stefan. "Beyond Benefits: Gratitude as a Response to Moral Regard." *Inquiry* 66, no. 10 (2023): 1821–40.

Rioux, Catherine. "Hope: Conceptual and Normative Issues." *Philosophy Compass* 16, no. 3 (2021): e12724.

Roberts, Robert C. "The Blessings of Gratitude: A Conceptual Analysis." In *The Psychology of Gratitude*, edited by Robert Emmons and Michael McCullough. Oxford University Press, 2004.

———. *Emotions: An Essay in Aid of Moral Psychology*. Cambridge University Press, 2003.

———. "The Normative and Empirical in the Study of Gratitude." *Res Philosophica* 92, no. 4 (2015): 883–914.

Roberts, Robert C., and Daniel Telech. "The Emotion-Virtue-Duty Triad of Gratitude: An Introduction to the Moral Psychology of Gratitude." In Roberts and Telech, *The Moral Psychology of Gratitude*.

———, eds. *The Moral Psychology of Gratitude*. Rowman and Littlefield, 2019.

Rosen, Gideon. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela Smith. Oxford University Press, 2015.

Rush, Michael. "Motivating Propositional Gratitude." *Philosophical Studies* 177, no. 5 (2020): 1191–211.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, and Blame*. Belknap Press, 2008.

Scheffler, Samuel. "Valuing." In *Equality and Tradition: Questions of Value in Moral and Political Philosophy*. Oxford University Press, 2010.

Schönherr, Julius. "Two Problems of Fitting Grief." *Analysis* 81, no. 2 (2021): 240–47.

Schultheiss, Oliver, and Martin Köllner. "Implicit Motives." In *Handbook of Personality: Theory and Research*, 4th ed., edited by Oliver John and Richard Robins. Guilford Press, 2008.

Searle, John. *Speech Acts*. Cambridge University Press, 1963.

Seligman, Martin E. P. *Authentic Happiness*. Free Press, 2002.

Setiya, Kieran. *Reasons Without Rationalism*. Princeton University Press, 2007.

Shoemaker, David. "Qualities of Will." *Social Philosophy and Policy* 30, nos. 1–2 (2013): 95–120.

———. *Responsibility from the Margins*. Oxford University Press, 2015.

Smith, Adam. *The Theory of Moral Sentiments*. Edited by David Daiches Raphael and Alec Lawrence Macfie. Oxford University Press, 1976.

Smith, Angela. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236–71.

Sommers, Tamler. "Partial Desert." In *Oxford Studies in Agency and Responsibility*, vol. 1, edited by David Shoemaker. Oxford University Press, 2013.

Story, Daniel. "Interpersonal Moral Luck and Normative Entanglement." *Ergo* 6, no. 21 (2019): 601–16.

Strawson, P. F. "Freedom and Resentment." In *Free Will*, 2nd ed., edited by Gary Watson. Oxford University Press, 2003.

Telech, Daniel. "Forgiveness and Moral Luck." In *Oxford Studies in Normative Ethics*, vol. 14, edited by Mark Timmons. Oxford University Press, 2024.

Telech, Daniel, and Hannah Tierney. "The Comparative Nonarbitrariness Norm of Blame." *Journal of Ethics and Social Philosophy* 16, no. 1 (2019): 25–43.

Tsang, Jo-Ann. "Gratitude and Prosocial Behaviour: An Experimental Test of

Gratitude." *Cognition and Emotion* 20, no. 1 (2006): 138–48.
Velleman, J. David. "Practical Reflection." *Philosophical Review* 94 (1985): 33–61.
———. "What Good Is a Will?" In *Action in Context*, edited by Anton Leist. De Gruyter, 2007.
Walker, A. D. M. "Gratitude and Gratefulness." *Proceedings of the Aristotelian Society* 81 (1980): 39–55.
Watkins, Phillip. *Gratitude and the Good Life: Toward a Psychology of Appreciation*. Springer, 2014.
Wedgwood, Ralph. "Instrumental Rationality." In *Oxford Studies in Metaethics*, vol. 6, edited by Russ Shafer-Landau. Oxford University Press, 2011.
Wood, Alex, Jeffrey Froh, and Adam Geraghty. "Gratitude and Well-Being: A Review and Theoretical Integration." *Clinical Psychology Review* 30, no. 7 (2010): 890–905.
Zagzebski, Linda. *Exemplarist Moral Theory*. Oxford University Press, 2017.
Zimmerman, Michael. "Taking Luck Seriously." *Journal of Philosophy* 99, no. 11 (2002): 553–76.

Journal of Ethics and Social Philosophy
VOL. 31, NO. 1 · FEBRUARY 2026
https://doi.org/10.26556/jesp.v31i1.4433

WHAT IS (FUNDAMENTALLY OR PER SE) WRONG WITH COLONIALISM

A REPLY TO AGRAWAL AND BUCHANAN

Daniel Weltman

WHAT is wrong with colonialism? Too much to mention. Much recent discussion of colonialism's wrongness focuses on a narrower question: Is there something per se wrong with colonialism, such that even colonialism absent all of the killing, enslaving, genocide, and so on would be wrong? Some argue that colonialism is per se wrong.[1] Others argue that colonialism is not per se wrong but is rather (as Margaret Moore puts it) only "contingently unjust."[2] Although in practice, it has always been wrong, in principle, we can imagine instances of colonialism that are not wrong.[3] This is an important question because (for instance) some wish to defend "civilizing" colonialism (either in principle or in practice) because it could promote justice.[4]

One recent analysis of colonialism by Ritwik Agrawal and Allen Buchanan bypasses discussion of colonialism's per se wrongness.[5] Agrawal and Buchanan instead raise a new topic: the "fundamental" wrong of colonialism. They argue that existing accounts of colonialism's wrongness do not accurately illuminate the fundamental wrong. The fundamental wrong is the assumption that those who are colonized are unequal because incapable: whereas most adults are

1 E.g., Altman and Wellman, *A Liberal Theory of International Justice*, 12–16; Ypi, "What's Wrong with Colonialism"; Stilz, "Decolonization and Self-Determination" and *Territorial Sovereignty*, 131–33; Bufacchi, "Colonialism, Injustice, and Arbitrariness"; van Wietmarschen, "The Colonized and the Wrong of Colonialism"; Renzo, "Why Colonialism Is Wrong"; Ferguson and Veneziani, "Territorial Rights and Colonial Wrongs"; and Nine, "Colonialism, Territory and Pre-Existing Obligations."

2 Moore, "Justice and Colonialism."

3 Valentini, "On the Distinctive Procedural Wrong of Colonialism"; and Weltman, "Colonialism, Injustices of the Past, and the Hole in Nine," "Colonialism Is Per Se Wrong Only if Colonialism Is Not Per Se Wrong," and "Saving Cosmopolitanism from Colonialism."

4 Weltman, "Saving Cosmopolitanism from Colonialism," 25.

5 Agrawal and Buchanan, "The Fundamental Wrong of Colonialism" (hereafter cited parenthetically).

regarded as autonomous, colonizers regard the colonized as similar to "children or the mentally deficient" (186).

We should reject Agrawal and Buchanan's arguments. They are wrong to charge their opponents with giving an incorrect account of the fundamental wrongness of colonialism. Their opponents are focused on the per se wrong of colonialism, and Agrawal and Buchanan change the topic. Moreover, their account of the fundamental wrong is not compelling because some instances of obviously wrongful colonialism do not instantiate the ostensibly fundamental wrong they propose.

1. FUNDAMENTAL VERSUS PER SE WRONG

Agrawal and Buchanan note earlier discussions of colonialism's per se wrongness (193), including Ypi's, Stilz's, and Renzo's discussions, which focus respectively on the form of association instantiated by colonialism, the inability of colonized peoples to affirm the colonial regime, and the violation of the self-determining agency of the colonized. But Agrawal and Buchanan change the terms of the debate and ask what the fundamental wrong of colonialism is, regardless of whether that fundamental wrong is essential to colonialism (and thus regardless of whether it renders colonialism per se wrong). The fundamental wrong is the "wrong from which the other serious wrongs derive and/or that is a violation of a more basic moral principle than is the case with other wrongs" (185).

Changing the debate is not necessarily objectionable. Where Agrawal and Buchanan err is in finding opponents among those who defend the per se wrongness of colonialism. They argue that Ypi, Stilz, and Renzo are charitably read as talking not about the per se wrong of colonialism but about "either the fundamental wrong of colonialism or … the most serious wrong among the wrongs that are grounded in or derivative upon the fundamental wrong" (189). Applying this ostensibly charitable gloss, they conclude that Ypi, Stilz, and Renzo all have bad accounts of the fundamental wrong of colonialism (190–92). But Ypi, Stilz, and Renzo have *no* account of the fundamental wrong of colonialism because they are not interested in the question, and it is not charitable to read them as being interested in this question. Ypi, Stilz, and Renzo are interested in the per se wrongness of colonialism. This means Agrawal and Buchanan's objections to Ypi, Stilz, and Renzo fail.

For instance, Agrawal and Buchanan argue that "views like that of Stilz, Renzo, and Ypi, so far as they focus [on] only the political dimension of colonialism, do not convey the full extent of colonial humiliation and the resulting, lasting damage to the self-conception of the (formerly) colonized" (191). The proper response is that the goal is not to convey this but rather to account for

colonialism's per se wrongness, which relies on a feature that may be only incidentally (or even not at all) related to these horrific results.

Ypi, for instance, agrees that "colonialism is wrong for many reasons," including the fact that it involves, among many other things, "burning native settlements, torturing innocents, slaughtering children, enslaving entire populations, exploiting the soil and natural resources available to them, and discriminating on grounds of ethnicity and race."[6] But she says that her goal is "to clarify what is wrong with colonialism, over and above these familiar outrages. Although an account focusing on the brutality of this practice would capture most of the wrong of colonialism … it would leave unchallenged more subtle forms of it."[7] She is "interested in what makes colonialism, even benign colonialism, wrong as such."[8] Thus, she is *very* explicit about *not* aiming to characterize any contingent wrong-making features of colonialism, no matter how bad those are and no matter how fundamental they are. She wants to focus on *only* what is per se wrong with colonialism, and this wrongness attaches even to benign colonialism.

Stilz, similarly, is concerned to give an account of colonialism's wrongness that would "rule out a *benign* colonial regime if it did a reasonable job at providing good governance."[9] Renzo too is explicit about looking for a wrong of colonialism "that cannot be reduced to the systemic perpetration of … more familiar crimes."[10] Renzo asks us to imagine "the benign form of government described by the apologists of colonialism" and suggests that "no matter how benevolent we imagine a particular master to be," we will still want to explain what is wrong with it by finding the per se wrong of colonialism.[11]

It should therefore be clear why Agrawal and Buchanan's focus on the fundamental wrong of colonialism is misplaced insofar as it is characterized as an argument against those who endorse the view that colonialism is per se wrong. The question of per se wrongness is separate from the question of fundamental wrongness. Thus, the objections that Agrawal and Buchanan level against Ypi, Renzo, and Stilz—like the claim that the wrong they identify is not more serious or more fundamental than the wrong Agrawal and Buchanan point to—fail (190–92). The objections target arguments that Ypi, Renzo, and Stilz do not make. Ypi, Renzo, and Stilz are not interested in the fundamental

6 Ypi, "What's Wrong with Colonialism," 162.

7 Ypi, "What's Wrong with Colonialism," 162.

8 Ypi, "What's Wrong with Colonialism," 161. I thank an anonymous reviewer for this journal for suggesting that I cite these statements Ypi makes.

9 Stilz, *Territorial Sovereignty*, 91.

10 Renzo, "Why Colonialism Is Wrong," 348.

11 Renzo, "Why Colonialism Is Wrong," 348.

wrong of colonialism, and their accounts cannot be fairly accused of giving a bad description of the fundamental wrong. Indeed, they all explicitly aim to adduce a wrong of colonialism that applies even in cases of benign colonialism, which would presumably not instantiate the fundamental wrong.

Agrawal and Buchanan consider something like this point and attempt to preclude it. They say that "we will proceed on the charitable assumption that by 'the wrong' of colonialism these authors do not mean a wrong unique to colonialism, since it is obvious that many non-colonial regimes fail to satisfy the criteria they advance" (189). Whatever the merits of this argument, we cannot therefore conclude that these authors must therefore be talking about the fundamental wrong of colonialism in the sense Agrawal and Buchanan have in mind. Agrawal and Buchanan present a false dichotomy. As they frame it, either we discuss the unique wrong of colonialism, or we discuss the fundamental wrong of colonialism. But there is a third option: we discuss the per se wrong of colonialism, whether or not that wrong is unique. Agrawal and Buchanan might be right that the wrong is not unique. But this does not mean that it is not a per se wrong or that it is a fundamental wrong.

Ypi, Renzo, and Stilz may or may not believe that the per se wrong of colonialism is also unique to colonialism, but this is a separate issue. Certainly it is not charitable to Ypi, Renzo, and Stilz to read them as addressing the fundamentality question if what they say about fundamentality is not convincing, as Agrawal and Buchanan claim. We should read Ypi, Renzo, and Stilz as doing what they *explicitly* say they are doing: we should read them as talking about the per se wrong of colonialism (and as not looking at all for what is fundamentally wrong with colonialism). If the *only* options were to read Ypi, Renzo, and Stilz as looking for the unique wrong *or* the fundamental wrong, Agrawal and Buchanan would have a point. But Ypi, Renzo, and Stilz may wrongly think that the per se wrong is a unique wrong, and thus we can reject what they say about whether the per se wrong of colonialism is unique but accept that they are describing a per se wrong, not a fundamental wrong.[12]

Agrawal and Buchanan might then respond that although it is true that their objections miss the mark, we should admit that because these opposing theories aim to answer a different question about per se wrongness rather than the question about fundamentality, these theories do not give a good

12 I think Ypi incorrectly thinks that the per se wrong of colonialism is a unique wrong, that Renzo probably thinks this given his discussion of the wrong as "distinctively wrong" ("Why Colonialism Is Wrong," 348), and that Stilz does not make this mistake. Valentini agrees with this reading of Ypi ("On the Distinctive Procedural Wrong of Colonialism"). My argument does not rely on the claim that Ypi, Renzo, or Stilz think that the per se wrong of colonialism is unique or on the claim that if any of them think it is unique, they are wrong to think this.

description of the fundamental wrong. This would give us a reason to prefer Agrawal and Buchanan's account of the fundamental wrong. We therefore turn to the more central point, which is what the fundamental wrong of colonialism is and whether Agrawal and Buchanan have accurately characterized it.

2. FUNDAMENTALITY

The fundamental wrong of colonialism, according to Agrawal and Buchanan, is "the non-autonomy assumption" (186). It is "the belief that colonized peoples were in effect not competent adults, that they were inferiors incapable of managing their own lives and therefore were suitable objects of control by superiors" (186).

As noted above, Agrawal and Buchanan characterize fundamentality as a property of a wrong such that it is the "wrong from which the other serious wrongs derive and/or that is a violation of a more basic moral principle than is the case with other wrongs" (185). It is hard to parse the definition because "and" and "or" mean different things, and conjoining them with a slash does not adequately elucidate the role that each word plays in the definition. I take the most charitable reading of Agrawal and Buchanan to be one according to which fundamentality consists of purported legitimation of other wrongs. The fundamental wrong of colonialism is the one that purportedly legitimates the most serious wrongs of colonialism—namely, economic exploitation and humiliation. This reading makes sense of almost all of the claims that Agrawal and Buchanan make about fundamentality and avoids issues with other possible interpretations of the idea.

Substantiating this reading of Agrawal and Buchanan is unnecessary for two reasons. First, Agrawal and Buchanan do not adequately elucidate what it means for a wrong to be fundamental so it is hard to say anything conclusive. Second and more importantly for the purposes of evaluating their account, we do not need to resolve the confusion. Any plausible reading of fundamentality is still subject to the objection raised below.

This objection is that there are many terrible instances of colonialism that do not instantiate the so-called fundamental wrong (or did so only some time after the colonization began). This is true whether we understand fundamentality in the way I have glossed it or in any other plausible way (like being a violation of a more basic moral principle). Whatever a fundamental wrong of colonialism is, we should expect to find it in typical and terrible instances of colonialism. This should make us suspect that Agrawal and Buchanan have not identified the fundamental wrong of colonialism. The objection can be illustrated with representative cases.

3. PROBLEM CASES FOR AGRAWAL AND BUCHANAN'S ACCOUNT

If the fundamental wrong of colonialism is an attitude that the colonized are not autonomous, this implies that none of the following cases instantiate the fundamental wrong of colonialism (at least in the initial stages of the colonization):

> *England and America*: King James I of England grants a charter to various English citizens, giving them the right to establish colonies in Virginia, specifying that the colonists must send one-fifth of the gold and silver and one-fifteenth of the copper they mine back to the king. The purposes of the colony include enriching the king and exerting British control over the Americas against competing European powers like Spain. The charter claims that colonization involves the "propagating of Christian religion to such People, as yet live in Darkness and miserable Ignorance of the true Knowledge and Worship of God" and characterizes the native inhabitants as "Infidels and Savages."[13] But the Indigenous Americans are not characterized as inferiors in the sense of being akin to children and unable to govern themselves at all, and indeed, the charter gives the colonists power to "establish and cause to be made a Coin . . . for the more Ease of Traffick and Bargaining between and amongst them and the Natives there."[14] The expectation is that the native inhabitants are rational actors capable of engaging in economic activity.
>
> *France and Algeria*: France invades Algeria "on the pretense of a diplomatic dispute revolving around French debts owed to the Bacri family and the Ottoman *dey* (head) of the Regency of Algiers that had resulted in the *dey* hitting the French consul in the face" but in fact for a variety of reasons, including King Charles X's desire to "present the French monarchy as the Christian defender of civilization against Muslim tyranny."[15] The assumption that the Algerians are childlike and not fully autonomous would, if anything, undercut the force of the insult that the blow represents and so cannot reasonably be adverted to as the fundamental wrong grounding other things France did.
>
> *Japan and Korea*: Japan colonizes Korea for a variety of reasons, including a desire to protect Japanese sovereignty by controlling surrounding areas and the expansion of Japan's role on the international stage. The view that the Koreans are incapable of ruling themselves is not a driving

13 Thorpe, *The Federal and State Constitutions*, 7:3784.

14 Thorpe, *The Federal and State Constitutions*, 7:3786.

15 Rabinovitch, "The Quality of Being French Versus the Quality of Being Jewish," 817.

> factor, and indeed, the Japanese colonial government clashes with existing Japanese settlers who disapprove of the colonial government's treatment of Koreans as potential Japanese citizens rather than as uncivilized barbarians.[16]

Similar cases, real or imagined, can be adduced. What they have in common is that they are awful (and paradigmatic) cases of colonialism yet are missing what Agrawal and Buchanan label the fundamental wrong of colonialism. Moreover, they all include economic exploitation, practices of humiliation, or both. Thus, they cannot be accused of missing Agrawal and Buchanan's point that any account of the fundamental wrong of colonialism should engage with the economic injustices and the humiliation involved in colonialism (186–89), which are indeed terrible. So Agrawal and Buchanan's account should be rejected.[17]

Ashoka University
danny.weltman@ashoka.edu.in

REFERENCES

Agrawal, Ritwik, and Allen Buchanan. "The Fundamental Wrong of Colonialism." *Philosophy and Public Affairs* 53, no. 2 (2025): 184–96.

Altman, Andrew, and Christopher Heath Wellman. *A Liberal Theory of International Justice*. Oxford University Press, 2009.

Bufacchi, Vittorio. "Colonialism, Injustice, and Arbitrariness." *Journal of Social Philosophy* 48, no. 2 (2017): 197–211.

Ferguson, Benjamin, and Roberto Veneziani. "Territorial Rights and Colonial Wrongs." *European Journal of Philosophy* 29, no. 2 (2021): 425–46.

Moore, Margaret. "Justice and Colonialism." *Philosophy Compass* 11, no. 8 (2016): 447–61.

Nine, Cara. "Colonialism, Territory and Pre-Existing Obligations." *Critical Review of International Social and Political Philosophy* 26, no. 2 (2023): 277–87.

Rabinovitch, Simon. "The Quality of Being French Versus the Quality of Being Jewish: Defining the Israelite in French Courts in Algeria and the Metropole." *Law and History Review* 36, no. 4 (2018): 811–46.

Renzo, Massimo. "Why Colonialism Is Wrong." *Current Legal Problems* 72, no. 1 (2019): 347–73.

16 Uchida, *Brokers of Empire*, 21.

17 For feedback on this discussion note, I thank two anonymous reviewers for this journal.

Stilz, Anna. "Decolonization and Self-Determination." *Social Philosophy and Policy* 32, no. 1 (2015): 1–24.

———. *Territorial Sovereignty: A Philosophical Exploration*. Oxford University Press, 2019.

Thorpe, Francis Newton, ed. *The Federal and State Constitutions, Colonial Charters, and Other Organic Laws of the States, Territories, and Colonies Now or Heretofore Forming the United States of America*. 7 vols. Government Printing Office, 1909.

Uchida, Jun. *Brokers of Empire: Japanese Settler Colonialism in Korea, 1876–1945*. Harvard University Asia Center, 2011.

Valentini, Laura. "On the Distinctive Procedural Wrong of Colonialism." *Philosophy and Public Affairs* 43, no. 4 (2015): 312–31.

Van Wietmarschen, Han. "The Colonized and the Wrong of Colonialism." *Thought* 7, no. 3 (2018): 170–78.

Weltman, Daniel. "Colonialism, Injustices of the Past, and the Hole in Nine." *Critical Review of International Social and Political Philosophy* 26, no. 2 (2023): 288–300.

———. "Colonialism Is *Per Se* Wrong Only if Colonialism Is Not *Per Se* Wrong: Supersession and the Bourgeois Predicament." *Public Affairs Quarterly* 38, no. 4 (2024): 239–66.

———. "Saving Cosmopolitanism from Colonialism." *Ethics and Global Politics* 17, no. 4 (2024): 25–44.

Ypi, Lea. "What's Wrong with Colonialism." *Philosophy and Public Affairs* 41, no. 2 (2013): 158–91.

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

Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

جامعة نيويورك أبوظبي
NYU ABU DHABI