

Georgia State University

ScholarWorks @ Georgia State University

Philosophy Theses

Department of Philosophy

8-2024

Playing the Game of Praise and Blame: On the Justification of our Responsibility Practices

Justice Cabantangan
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Cabantangan, Justice, "Playing the Game of Praise and Blame: On the Justification of our Responsibility Practices." Thesis, Georgia State University, 2024.

doi: <https://doi.org/10.57709/37088731>

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

Playing the Game of Praise and Blame: On the Justification of our Responsibility Practices

by

Justice Cabantangan

Under the Direction of Eddy Nahmias, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2024

ABSTRACT

Moderate revisionists about moral responsibility, such as Manuel Vargas, attempt to preserve the backward-looking norms that ordinarily guide praise and blame while arguing that these norms are justified because of their effects. However, some doubt that backward-looking norms are good at achieving these effects; this is the objection from inefficacy. In this paper, I introduce an alternative to Vargas' view that holds that our responsibility practices are a game, as analyzed by Bernard Suits. I argue that this alternative avoids the objection from inefficacy while preserving attractive features of Vargas' view. Further, I argue that the game view yields benefits over other revisionist views to position the view as a promising account for the justification of our responsibility practices. If our responsibility practices are a game, then it affords the game view a valuable conceptual distinction unavailable on other views and ensures that praise and blame are attributed wholeheartedly.

INDEX WORDS: Moral responsibility, Free will, Revisionism

Playing the Game of Praise and Blame: On the Justification of our Responsibility Practices

by

Justice Cabantangan

Committee Chair: Eddy Nahmias

Committee: Andrew I. Cohen

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2024

TABLE OF CONTENTS

1	INTRODUCTION.....	1
2	TWO JUSTIFICATIONS FOR PRAISE AND BLAME.....	3
3	CHARTING A MIDDLE PATH.....	8
4	THE GAME OF PRAISE AND BLAME	13
5	WHY PLAY THE GAME?.....	20
6	A RESPONSE TO AN OBJECTION AND A QUALIFICATION.....	34
7	CONCLUDING REMARKS	37
8	BIBLIOGRAPHY	38

1 INTRODUCTION

Manuel Vargas' theory of moral responsibility walks a fine line between two disparate notions of responsibility. On one side is our ordinary, "backward-looking" notion of moral responsibility whereby praise and blame are justified in light of backward-looking considerations such as whether a blameworthy or praiseworthy action *was* actually performed, whether it *was* intentional, and so on. On the other side is a revisionist, "forward-looking" notion of responsibility whereby praise and blame are justified in light of forward-looking considerations such as the effects that praise or blame *will* have on its target, or the more global effects that an instance praise or blame *will* produce. Vargas' mission in taking the middle path is twofold: conserve the intuitive, backward-looking character of our responsibility practices *but also* justify our practices by appealing to their overall future effects, thereby avoiding reference to the controversial capacity for control commonly associated with our backward-looking responsibility practices.¹ However, in taking the middle path, a problem arises from each side of the line Vargas walks. In Section 2, I introduce the backward-looking and forward-looking notions of moral responsibility, and present problems for each of them that motivate the need for a middle path. In Section 3, I summarize one such middle path, proposed by Manuel Vargas, and lay out his response to the problems presented in Section 2. Then, I introduce a problem that arises for Vargas – the objection from inefficacy. In Section 4, I lay out an alternative account of responsibility that, I argue, can walk the middle path while avoiding the objection from inefficacy. If our responsibility practices are a game – assuming Bernard Suits' analysis of games – I argue that the objection from inefficacy loses its force. Further, I argue that such an

¹ Manuel Vargas, *Building Better Beings*, (New York: Oxford University Press, 2013), 165; P.F. Strawson, "Freedom and Resentment," in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, ed. Laura Waddell Ekstrom (Boulder: Westview, 2000), 203.

account can preserve Vargas' responses to the problems presented in Section 2. In Section 5, I further motivate this account of our responsibility practices by arguing that there are benefits to adopting the view that they are a game. Finally, I respond to an objection, introduce qualifications for my arguments, and offer closing remarks in Sections 6 and 7.

2 TWO JUSTIFICATIONS FOR PRAISE AND BLAME

In ordinary cases, when we praise or blame someone, we believe we are justified in doing so because of some facts about the past.² For instance, if my friend steals money from my unattended wallet, I believe I am justified in blaming them if they actually and intentionally did so. I consult these “backward-looking” considerations and arrive at the conclusion that blame is warranted towards my friend. This ordinary idea, however, runs into some trouble when we consider more closely the reasons why praise or blame is warranted in light of these backward-looking considerations.

The ordinary, backward-looking notion of responsibility seems to assume a distinctive kind of control – free will – that agents have over their actions such that it makes sense to hold them responsible and praise or blame them. This kind of control has often been cashed out as an ability to do otherwise or an ability to act in such a way that the action was “up to you”. However, it is naturalistically implausible that agents are capable of this kind of control. This is because virtually every account that attempts to flesh out what the ability consists in seems to lack “truth-relevant considerations that speak in its favor when it postulates requirements that exceed the known facts or the widely accepted ontologies of our current scientific understanding”.³ In other words, such accounts – libertarian accounts – end up proposing that free will consists in certain events or powers that, while possible under the best scientific theories, have no positive scientific evidence that vouch for the actual occurrence of such events or powers. For example, Robert Kane’s libertarian account suggests that we can act in a way that is “up to us” in moments when we have competing motivations which cause a “‘stirring up of chaos’ in the brain that makes it

² Here, I employ the same “broad” yet “familiar” sense of “justified” that Vargas does. Activities that are justified are those “that make good sense, that we have reason to be engaged in, or that are well supported by our various interests and cohere with our not-indefensible commitments”; Vargas, *Building Better Beings*, 158-9.

³ Vargas, *Building Better Beings*, 59.

sensitive to micro-indeterminacies at the neuronal level”.⁴ Although there is no evidence that definitively rules out the occurrence of such phenomena, as Vargas notes, “nothing in the brain sciences would lead an independent observer to conclude” that such phenomena actually occur.⁵ Notably, there are other skeptical arguments that cite complaints other than naturalistic implausibility.⁶ However, here, I only wish to maintain that the libertarian justification for our responsibility practices is naturalistically implausible. Notably absent from this discussion is talk of “compatibilist” theories – theories that attempt to flesh out a more naturalistically plausible free will in Vargas’ lights. I omit compatibilist theories from my discussion because I agree with Vargas that such species of free will would fall short from capturing the ordinary backward-looking notion of responsibility.⁷ Thus, if we insist that backward-looking considerations justify praise and blame and that they require libertarian powers, it becomes naturalistically implausible that agents can be held responsible for their actions. Given this difficulty, some “forward-looking” theorists have looked elsewhere for a more stable justification for our responsibility practices.

Forward-looking theorists of responsibility hold that our responsibility practices are justified by their effects. Some might hold that our responsibility practices are justified insofar as they “influence agents in socially desirable ways”.⁸ For instance, J. J. C. Smart holds that praise may be justified in light of fulfilling one of its functions: “encourag[ing] people to do actions of [a certain] class”.⁹ On such theories, if my friend steals money from my unattended wallet, I

⁴ John Martin Fischer et al., *Four Views on Free Will* (Malden (Ma) ; Oxford ; Victoria: Blackwell, 2014), 26.

⁵ Vargas, *Building Better Beings*, 62.

⁶ For a review of skeptical arguments see Gregg Caruso and Derk Pereboom, *Moral Responsibility Reconsidered* (Cambridge University Press, 2022), 20-36.

⁷ Vargas, *Building Better Beings*, 46.

⁸ Manuel Vargas, “Moral Influence, Moral Responsibility,” in *Essays on Free Will and Moral Responsibility*, ed. Nick Trakakis and Daniel Cohen (Newcastle: Cambridge Scholars Press, 2008), 91.

⁹ J. J. C. Smart, “Free Will, Praise and Blame,” *Mind* 70, no. 279 (1961): 304.

would be justified in blaming them to the extent that blaming them would yield certain effects – e.g. deterring my friend from stealing in the future. In contrast to the backward-looking notion discussed above, I consult forward-looking considerations – whether my blame *will* yield certain effects – and arrive at the conclusion that blame is warranted. To be clear, forward-looking theories may differ in detail as to which forward-looking considerations should count as justifications for praise and blame, but the relevant feature to highlight here is one shared by all such theories: they justify our responsibility practices by appeal to their effects. It is important to note that these theories may still require specific capacities to be considered a responsible agent, but the capacities that are required for responsible agency are generally taken to be more naturalistically plausible than the kind of control assumed by backward-looking notions of responsibility (i.e. the capacity to respond to reasons). However, forward-looking theories seem to conflict with our ordinary idea of moral responsibility and are thereby considered revisionist – they claim that our ordinary idea of moral responsibility ought to be revised. Thus, although a forward-looking theory may indeed yield less debatable justificatory grounds for our responsibility practices than its backward-looking counterpart, such theories run into a problem of their own in light of their revisionism.

The objection can be put in several ways. For Kelly McCormick, “this problem can be understood as the worry that there is no guarantee that a revisionist prescriptive account of ‘moral responsibility’ will in fact be a genuine account of responsibility”.¹⁰ If revisionists admit that the “moral responsibility” they speak of deviates from the ordinary concept of moral responsibility, then it is unclear whether this revisionist “moral responsibility” is one that we care about. For Daniel Coren, a “pronounced asymmetry between the blame and the reason for

¹⁰ Kelly McCormick, “Anchoring a Revisionist Account of Moral Responsibility,” *Journal of Ethics and Social Philosophy* 7, no. 3 (Dec. 2013): 5.

the blame” arises from forward-looking responsibility theories when applied to paradigmatic instances of extremely blameworthy action.¹¹ Even in blameworthy cases where extreme wrongs are perpetrated, such as massacres or other large-scale harms, forward-looking theories seem to commit us to the unintuitive position that blame is not warranted because and in proportion to such wrongs. For John Doris, forward-looking theories of responsibility threaten eliminativism.¹² Such theories are at risk of being so far removed from the ordinary notion that they seem to eliminate rather than revise our existing responsibility practices. Let us call this family of objections, the objections from misdescription, as they claim that the problem with forward-looking theories is their failure to map onto our ordinary responsibility practices.

So, each theory faces a problem. Backward-looking responsibility practices assume that we have a naturalistically implausible capacity for control over our actions, while the revisionist nature of forward-looking responsibility practices raises several worries about whether the “moral responsibility” under such theories is moral responsibility at all. This discussion illuminates the desiderata for a “middle path” between these two notions of moral responsibility:

D1. Avoid justifying our responsibility practices by appeal to an implausible capacity for control.

D2. Avoid the objections from misdescription; do not stray too far from our ordinary notion of moral responsibility.

“Moderate revisionists” such as Manuel Vargas aim to fulfill these desiderata with theories of responsibility that are sufficiently conservative relative to the ordinary notion while looking

¹¹ Daniel Coren, “Moral Responsibility Must Look Back,” *American Philosophical Quarterly*, (forthcoming).

¹² John Doris, “Doing without (arguing about) desert,” *Philosophical Studies* 172 (2015): 2630.

elsewhere for a justification for responsibility practices. In the next section, I turn to Vargas' theory and explain how his theory fulfills D1 and, more importantly, D2, but then I introduce an important problem for it.

3 CHARTING A MIDDLE PATH

Vargas is similar to forward-looking theorists in proposing that praise and blame are justified by their effects, namely, “[producing] agents that, over time and in a wide range of contexts, are suitably responsive to moral considerations” (or just “building better beings”).¹³ Thus, also like other forward-looking theorists, Vargas fulfills *D1*. He avoids appealing to a naturalistically implausible capacity for control to justify our responsibility practices by instead appealing to an effect of our practices: building better beings.

Unlike other forward-looking theories, however, Vargas argues that his account has the resources to respond to the objection from misdescription and thereby fulfill *D2* by preserving the backward-looking norms that ordinarily guide our practice. Vargas and Doris observe that our ordinary responsibility practices are guided by backward-looking norms that are best described under a “quality-of-will” account of responsibility norms.¹⁴ Under the quality-of-will account, responsibility norms revolve around the principle that “when one acts with a bad will, one is blameworthy and where one acts with a good will, one is praiseworthy”.¹⁵ Quality-of-will norms tell us to look back at the features of some morally-relevant action in order to evaluate the actor’s quality of will. To take our previous example, when I blame my friend because she actually and intentionally stole money from my wallet, I am justified in doing so because her actually and intentionally stealing indicates her acting with a bad or poor quality of will. If my friend had only accidentally done so, then I am not justified in blaming my friend (at least to the same extent as in the former case) because the action was not indicative of a poor quality of will. It is important to recognize the scope of these norms. Responsibility norms quantify only over

¹³ Vargas, *Building Better Beings*, 177.

¹⁴ Vargas, *Building Better Beings*, 192; Doris, “Doing without (arguing about) desert,” 2626.

¹⁵ Vargas, *Building Better Beings*, 160.

instances of praise and blame, rather than over actions more generally. These norms give us insight only with regards to whether some instance of praise or blame is “fitting” or “correct” but are silent on issues about what is all-things-considered “good” to do. For example, while blaming the hostage taker is fitting, it may not be all-things-considered good to blame the hostage taker while they are currently holding hostages. Relatedly, quality-of-will norms do not specify what consists in a good or bad quality of will; they only specify how praise and blame ought to be distributed given one’s quality of will. Specifying what consists in a good or bad quality of will is the work of a moral theory or theory of practical reason. Further, I admittedly use the terms “praise” and “blame” imprecisely both here and throughout; what “praise and blame” amounts to turns on what the correct account of praise and blame is.¹⁶ However, here, I need not assume a certain account of praise and blame; I only posit that whatever praise and blame turn out to be, they are ordinarily guided by these backward-looking, quality-of-will norms.

Vargas’ moderately revisionist account can preserve these quality-of-will norms characteristic of our ordinary responsibility practices – despite holding that praise and blame are ultimately justified by their effects – by proposing that our practices have a two-tier structure of justification. On his account, particular instances of the practice of moral responsibility (e.g. specific instances of praising or blaming) are guided by these quality-of-will norms and are, thus, justified by backward-looking considerations. However, these norms are justified in light of forward-looking considerations – the whole system of responsibility norms is justified by the effects it will yield given our adherence to it. At the first “tier”, our ordinary, backward-looking considerations (per quality-of-will norms) justify a particular instance of attributing moral

¹⁶ For a review on attempts to characterize blame in particular, and a promising, original theory of blame see David Shoemaker and Manuel Vargas, “Moral Torch Fishing: A Signaling Theory of Blame,” *Noûs* 55, no. 3 (October 24, 2019): 581–602.

responsibility. For example, I can still justify my blame towards my friend for stealing money from my wallet because she actually and intentionally stole. At the higher-order second “tier”, we can *look forward* to the effects that Vargas thinks ultimately justify our backward-looking responsibility norms: building better beings. That my blame adheres to backward-looking norms is justified because it contributes to building better beings. While a particular instance of moral responsibility is justified by backward-looking norms, this system of norms is justified by forward-looking considerations. Vargas’s revisionism can thus be moderately conservative in the face of the objections from misdescription. By preserving the backward-looking norms, Vargas can claim that his account accommodates our ordinary responsibility practices and thereby fulfills D2. What makes Vargas’ theory revisionist is only “that it conflicts with beliefs we may have about the reason to be [backward-looking]”.¹⁷ Because such beliefs are not as widely held as the intuition that instances of praise and blame are justified in light of backward-looking considerations, a theory of responsibility that conflicts with these beliefs is not necessarily one that strays far from our ordinary notion of moral responsibility.

However, in holding that our backward-looking norms of responsibility are justified in light of forward-looking considerations, Vargas gets caught up in a problem raised by more radically revisionist forward-looking theorists of responsibility. For instance, Kelly McCormick points out that “if there are in fact alternative systems capable of [building better beings], then our continued participation in the [backward-looking] responsibility system is not the only way to promote [such] agency”.¹⁸ Even if we grant that our backward-looking practices are justified in light of their effects, it seems that any other set of practices that achieve (or better achieve) the same effects is at least equally (or more) justified. Derk Pereboom, for instance, argues that

¹⁷ Vargas, “Desert, responsibility, and justification,” 2668.

¹⁸ McCormick, “Anchoring a Revisionist Account of Moral Responsibility,” 15.

conserving the backward-looking norms, as Vargas does, yields a responsibility practice that is “insensitive” to the forward-looking goal of building better beings.¹⁹ He further argues that adopting a more radically revisionist “single-tier” theory – whereby forward-looking considerations *do* factor into particular instances of holding responsible – would yield more efficacious responsibility practices.²⁰ Victoria McGeer, relatedly, points out that the conserved backward-looking norms of Vargas’ theory begin to look inefficacious, or at best arbitrary, at achieving the supposed goals of our responsibility practices once we realize that these norms usually warrant responsibility ascriptions towards special agents (such as psychopaths) that cannot be influenced in such a way that would contribute towards building better beings.²¹ Because these objections claim that Vargas’ conservatism yields a less efficacious set of responsibility practices than viable forward-looking practices, let us refer to these objections as the objections from inefficacy.

Avoiding the objections from inefficacy expands the list of desiderata beyond *D1* and *D2* for charting a middle path between backward-looking and forward-looking notions of responsibility:

D3. Avoid the objections from inefficacy; either show that responsibility practices under the account are efficacious or that the account has resources to admit its inefficacy without being problematic.

¹⁹ Derk Pereboom, “Undivided Forward-Looking Moral Responsibility,” *The Monist* 104 (2021): 484-497.

²⁰ Vargas thinks we do not need to seek out and adopt the most effective set of responsibility practices to build better beings if a more effective set would incur substantial costs. For example, it is a cost for any set of responsibility practices if it is not easily able to be internalized given our moral psychology. However, Pereboom argues that a single-tier theory would, all-things-considered, still be a better option than Vargas’ theory by appealing to studies in the social psychology literature.

²¹ Victoria McGeer, “Building a better theory of responsibility,” *Philosophical Studies* 172 (2015): 2639-2640.

Vargas might argue that his account can satisfy *D3* by either arguing that a backward-looking practice is more efficient than a forward-looking practice at achieving the justifying effects or arguing that any radically forward-looking practice fails to respond to the objection from misdescription. However, the radically forward-looking theorist might just respond to Vargas by arguing that a forward-looking practice is more efficient than a backward-looking practice and/or the benefits afforded by the efficacy of forward-looking practices outweigh the costs alleged by the objection from misdescription. For instance, the threats of eliminativism or of describing a subject that is not of interest to us when we talk about “responsibility” may be trade-offs that the radically forward-looking theorist is willing to make in exchange for greater efficacy relative to Vargas’ account. In what follows, I explore a different middle path, an alternative moderately revisionist view that avoids the argumentative stalemate over the efficacy of backward-looking and forward-looking responsibility practices, and the relative weights of efficacy and conservatism. If our responsibility practices are a special kind of activity – a game – then, contra Vargas and Pereboom, its inefficacy at bringing about its goals would not necessarily count as a strike against it. Along with fulfilling *D3*, I will also argue that such an account, what I will call *the game view*, can preserve an important feature of Vargas’ account to fulfill *D2*. I leave *D1* to be addressed in the following section, where I address the game view’s benefits.

4 THE GAME OF PRAISE AND BLAME

The comparison between our responsibility practices and games is not unfamiliar to responsibility theorists; it is typically employed to illustrate *some* feature of our responsibility practices that happens to resemble a feature of games. Vargas, in fact, argues that engaging in responsibility practices as described under his view is similar to engaging in a sport.²² In this section, I take the similarity between our responsibility practices and games a step further. I will introduce the game view of praise and blame, which holds that our practices of attributing responsibility *are* a game, assuming Bernard Suits' analysis of games. I will map what Suits takes to be constitutive of games onto our responsibility practices, laying out what our responsibility practices could look like under the game view – effectively describing a responsibility game. Further, I will argue that such a view would fulfill *D2* and *D3*.

For Suits, an activity is a game if and only if it is an “(a) activity directed towards bringing about a specific state of affairs, (b) using only means permitted by rules, (c) where the rules prohibit more efficient in favour of less efficient means, and (d) where such rules are accepted just because they make possible such activity”.²³ Consider the 100m dash for example. It is (a) an activity directed towards a specific state of affairs – reaching the finish line before the other competitors. This state of affairs constitutes what Suits would call the 100m dash's “prelusory goal” – the state of affairs that a game is directed towards but described without reference to the rules of the game. Notice, however, that I can achieve this prelusory goal by riding my motorcycle down the track, but in doing so, I would not be playing the game. To play a game, one must pursue its prelusory goal while *also* (b) adhering to its rules – staying on the

²² Vargas, “Desert, responsibility, and justification,” 2667.

²³ Bernard Suits, *The Grasshopper: Games, Life, and Utopia* (Toronto: University of Toronto Press, 1978), 34; Bernard Suits, “What Is a Game?,” *Philosophy of Science* 34, no. 2 (1967): 156. Parentheses mine.

track, starting only after the starting gun, etc. In other words, we must pursue the “lusory goal”: reaching the finish line before the other competitors *while* following the rules of the game. Further, (c) a game’s rules prohibit more efficient means towards achieving the prelusory goal. In the case of the 100m dash, the rules prohibit me from riding my motorcycle, using weapons against my competitors, etc.— means which would have been more efficient at achieving the prelusory goal than the means allowed by the rules. Lastly, what motivates one to accept these rules determines whether the activity that one is engaged in is a game. If you race through the length of the track *only* to defuse a bomb located past the finish line, then you are not *really* engaged in the 100m dash. Although you are coincidentally following the rules, you are not a genuine gameplayer. Genuine gameplayers must exhibit what Suits calls the “lusory attitude” – a distinctive motivational attitude towards the rules of a rule-guided activity. To demonstrate the lusory attitude one must (d) accept the rules because doing so makes participation in the 100m dash possible.

On the game view, our responsibility practices are first and foremost (a) directed towards bringing about a specific state of affairs. Under Vargas’ theory, the goal of our responsibility practices is to build better beings. I likewise hold that this is the state of affairs that our responsibility practices are directed towards. Thus, on the game view, the prelusory goal of our responsibility practices is to build better beings.

Further, on the game view, (b) the means of pursuing the responsibility game’s prelusory goal must be constrained by its rules. On Suits’ analysis, game rules are “proscriptions of certain means useful in achieving prelusory goals”.²⁴ Our ordinary, quality-of-will norms proscribe certain means useful in building better beings. Being social creatures, social rewards such as

²⁴ Suits, *The Grasshopper*, 37.

praise are just about, if not more, powerful than non-social rewards (e.g. food, money, etc.) in reinforcing certain behaviors. Evidence from neuroscience even suggests that the similarity between social and non-social rewards appear at the neurological level as “striatum activity correlates with valued social experiences in a similar manner to nonsocial rewards”.²⁵ Similarly, it is plausible that social punishments such as blame will have the same effects as non-social punishments, guiding targets of punishment away from punished behaviors.²⁶ Thus, there will be cases where praise or blame might be expressed towards agents who do not exhibit a good quality of will, but having recognized that they are praised for their conduct for moral reasons such agents would have moral reasons to repeat their conduct in the future. There may also be cases where praise and blame can foster the moral considerations-responsiveness of agents who are not the direct targets of praise or blame – spectators of instances of praise and blame. David Shoemaker and Vargas argue that a promising way to make sense of blame is as a signal about what one is morally committed to or what moral demands one has of others.²⁷ Note, however, that blame can serve its function as a signal even when directed towards agents who did not exhibit a poor quality of will. An agent may have only accidentally taken money from my wallet (having mistaken the wallet for hers), but this does not hinder the ability for my blame to signal to onlookers that I demand that my property rights be respected. Thus, if Shoemaker and Vargas’ characterization of blame is right and assuming that blame’s signal is taken to heart by others, then we have yet more cases in which holding an agent responsible would be conducive to building better beings despite the agent not exhibiting the appropriate quality of will.

²⁵ Jamil P. Bhanji and Mauricio R. Delgado, “The Social Brain and Reward: Social Information Processing in the Human Striatum,” *Wiley Interdisciplinary Reviews: Cognitive Science* 5, no. 1 (October 8, 2013), 63.

²⁶ Philip Jean-Richard-Dit-Bressel, Simon Killcross, and Gavan P. McNally, “Behavioral and Neurobiological Mechanisms of Punishment: Implications for Psychiatric Disorders,” *Neuropsychopharmacology* 43, no. 8 (July 1, 2018), 1639.

²⁷ Shoemaker and Vargas, “Moral Torch Fishing,” 587.

If quality-of-will norms were applied to the types of cases described above, they would proscribe means that are useful to building better beings. If this is so and building better beings constitutes the prelusory goal of the responsibility game, then the quality-of-will norms can constitute the responsibility game's rules. Rules proscribe more efficient in favor of less efficient means to achieving a game's prelusory goal. Under quality-of-will rules, "moves" within the responsibility game would be "invalid" insofar as they inaccurately track their target's quality of will – one would be breaking the rules by praising or blaming someone who clearly did not exhibit the appropriate quality of will even if breaking the rules would be conducive (or more conducive than following the rules) to building better beings.

The foregoing discussion reveals that the quality-of-will norms characteristic of our ordinary responsibility practices can be adapted into the rules of a responsibility game given Suits' definition of rules. If we adopt these quality-of-will rules, the game view can preserve Vargas' response to the objection from misdescription. To see how, recall that Vargas avoids the objection from misdescription and fulfills *D2* by making the forward-looking justifying effects relevant only at the second tier; at the first tier, Vargas preserves the application of quality-of-will norms. Despite the responsibility norms being justified by forward-looking considerations, it is the preservation that occurs the first tier – that particular instances of praise and blame are still justified by backward-looking considerations – that allows Vargas to deny the claim that his account is too far removed from our ordinary practices. If the game view adopts quality-of-will rules, particular instances of praise or blame are likewise justified by backward-looking considerations. Like our quality-of-will norms, the quality-of-will rules tell us to look back at the features of some morally-relevant action in order to evaluate the actor's quality of will. By adopting quality-of-will rules, the game view would hold that instances of praise or blame are

justified by backward-looking considerations much like Vargas' account. Thus, if the game view adopts such rules, it can preserve the conservative character of Vargas' account to avoid the objection from misdescription and thus fulfill *D2*. To be clear, any set of restrictions on efficient means of pursuing the building of better beings can constitute the rules of the responsibility game. However, quality-of-will rules yield a rather straightforward answer to *D2* (à la Vargas). Further, recall that the quality-of-will rules are based off of the norms already characteristic of our practices, saving us from needing to innovate by producing a set of rules unfamiliar to us. This gives us good reason to think that quality-of-will rules would make a good set of rules for a responsibility game. I will address the possibility of different sets of rules in Section 6.

So far, the game view nicely maps onto Vargas' account. The goal of our responsibility practices and the conservation of our backward-looking quality-of-will norms are features from Vargas' account that the game view can preserve, leveraging the latter to fulfill *D2*. However, the final two features of games illuminate important deviations from Vargas' view. For instance, if our responsibility practices are a game guided by backward-looking rules, then we can concede to Pereboom and McGeer that, contra Vargas, (c) the rules of the responsibility game prohibit more efficient in favor of less efficient means of achieving the prelusory goal. Riding my motorcycle or using weapons are certainly means that are more efficient at getting me to the finish line before my competitors than the means that the rules of the 100m dash deem valid. Similarly, as demonstrated above, there are instances in which breaking the quality-of-will rules would be more efficient at building better beings than the means the quality-of-will rules deem valid on the game view. What follows is that there are some practices that are more efficient at achieving the prelusory goal of the responsibility game than the responsibility game itself. Thus,

it is possible on the game view, as Pereboom insists, that a forward-looking practice is more efficient at building better beings.

However, even if we assume that a forward-looking practice is more efficacious at building better beings than the responsibility game, this need not leave the game view prone to the objection from misdescription and thus unable to fulfill *D3*. On the game view, (d) the rules of the responsibility game are adhered to because they make possible the activity of holding responsible; this illuminates a response to the objection from inefficacy. For Suits, the possibility of gameplay is the *only* necessary and sufficient reason to follow the rules of the game. I am justified in claiming that the reason why I do not ride my motorcycle on the track is *just that* I want to be able to participate in the 100m dash.²⁸ Thus, on the game view, adherence to the rules of the responsibility game is not necessarily justified in proportion to their efficacy at building better beings. The objection from inefficacy argues that a responsibility practice guided by only backward-looking norms is less justified because such a practice is less efficient at achieving its supposed justifying effects relative to a practice guided by forward-looking norms. The objection assumes that following a set of norms is justified if and only if doing so would be conducive to certain effects. Thus, it assumes that the possibility of participating in responsibility practices cannot be the only necessary and sufficient reason to follow a set of norms. However, if neither the achievement of the prelusory goal nor greater relative efficacy at achieving the prelusory goal is necessary to justify our participation in a game, then a different practice is not necessarily justified in proportion to its efficacy at bringing about that game's prelusory goal. That the use of a motorcycle affords me more efficient means of reaching the finish line than the means deemed valid by the rules of the 100m dash does not entail that the use of a motorcycle is more justified.

²⁸ Suits, *The Grasshopper*, 144.

Accordingly, if forward-looking practices allow us to build better beings more efficiently than backward-looking practices, it would not entail that forward-looking practices are more justified. The game view avoids the objection from inefficacy and thereby satisfies *D3* by denying the objection's assumption that efficacy is both a necessary and sufficient reason to follow a particular set of responsibility norms. If our responsibility practices are a game, then one is justified in adhering to the rules of the responsibility game just because it allows them to praise and blame.

5 WHY PLAY THE GAME?

In the last section, I argued that the game view, like Vargas' account, satisfies *D2* and, unlike Vargas' account, satisfies *D3*. If the backward-looking norms constitute the rules of the responsibility game, the game view can preserve the backward-looking character of our ordinary responsibility practices. Further, because our responsibility practices are a game, adhering to its rules is not justified by the efficacy of the means deemed valid by its rules at achieving its prelusory goal. Notably, however, my arguments depend on the important assumption of the game view that our responsibility practices are a game. Thus, the following concern arises: while it is one thing to say that the game view could satisfy both *D2* and *D3*, it is another thing to say that the game view is true.

My response to this concern could be called prescriptive rather than descriptive. I will not argue that our responsibility practices are, in fact, a game, or that deeper reflection will reveal that our responsibility practices are best described as a game. I will argue, rather, that our responsibility practices *should* be described as per the game view by arguing for the view's merits relative to other revisionist accounts. This approach is similar to Vargas' and other revisionist theorists insofar as I take the game view to be outlining "a picture of the normative anchors for our responsibility-characteristic practices, a picture that (ideally) also provides resources for illuminating the internal structure of the responsibility system".²⁹ Again, Vargas takes his account to be revisionist only insofar "that it conflicts with beliefs we may have about the reason to be [backward-looking]".³⁰ Similarly, my account may conflict only with beliefs we have about the reason to be backward-looking; any revision done would be manifest at the level of conceptual bookkeeping and may not manifest at all in our first-order practices. Much like

²⁹ Vargas, *Building Better Beings*, 159.

³⁰ Vargas, "Desert, responsibility, and justification," 2668.

other revisionist projects, the goal of the game view is to provide an adequate justification for our responsibility practices such that we have good reason to praise and blame in the way we do. As Vargas puts it, the goal of such theories “is not whether the full package of our conceptual commitments *about* responsibility can be defended”.³¹ Thus, the success of revisionist accounts does not rely on how well it conforms to our intuitions about what moral responsibility is at the second-tier. Although, as we have seen, sufficiently far deviations from our ordinary, first-tier responsibility practices may leave an account prone to the objection from misdescription and thereby at risk of leaving *D2* unsatisfied. However, the success of such accounts is primarily judged on the basis of theoretical considerations such as its justificatory adequacy, internal coherence, naturalistic plausibility, moral permissibility, etc.³²

In this section, I argue that the game view has theoretical benefits over and above the moderately revisionist view of Vargas and more radically revisionist views such as Smart and Pereboom who both hold that instances of praise and blame are directly justified by their effects. Namely, I argue that there are reasons to play the responsibility game over and above engaging in the responsibility practices as described by each of these two camps. Given this approach, I will also be arguing that the game view would satisfy *DI*; if we have good reasons to play the responsibility game, then we have a justification for praise and blame independent of the debate over the control condition. Importantly, the view’s merits do not make appeal to the responsibility game’s efficacy at achieving its prelusory goal; again, a game’s efficacy at achieving its prelusory goal does not justify participation in them. Thus, I appeal to other normative considerations that favor the game view.

³¹ Vargas, “Desert, responsibility, and justification,” 2661. Italics mine.

³² See McCormick, “Anchoring a Revisionist Account of Moral Responsibility,” 2.

5.1. The Radical Instrumentalist

The game view allows us to distinguish between genuine gameplayers and a kind of pseudo-gameplayer – the “radical instrumentalist”. This distinction illuminates a conceptual benefit of the game view; it allows the game view to make a valuable distinction unavailable on other revisionist views. A weakness of forward-looking theories is that they accommodate for radical instrumentalists in their responsibility practices – these are folks who issue praise/blame solely motivated by forward-looking concerns. Much like a radical instrumentalist of the 100m dash who runs the length of the track *only* to defuse a bomb located past the finish line, a radical instrumentalist of the responsibility game issues praise and blame *only* for the sake of achieving some effect(s). I claim that, for radical instrumentalists, praise and blame “come from the wrong place”. This claim echoes P.F. Strawson’s analysis of what he calls the “objective attitude”.³³ To hold the objective attitude towards someone, P.F. Strawson writes, is to see them “as an object of social policy, as a subject for what, in a wide range of sense, might be called treatment” or as something “to be managed or handled or cured or trained”.³⁴ That is, in holding the objective attitude towards someone, we view them solely in terms of what can be done with them; our interactions with them are only motivated by some effect or outcome. For radical instrumentalists of responsibility, targets of praise and blame are viewed merely in terms of what can be done with them; praise and blame are motivated by achieving some effect or outcome (i.e. successful treatment, management, etc.). Strawson observes that engagement in “inter-personal human relationships” with others is precluded in proportion to the extent that the objective attitude is held towards them.³⁵ Thus, radical instrumentalists, in praising and blaming as they

³³ Strawson, “Freedom and Resentment,” 190.

³⁴ Ibid.

³⁵ Ibid.

do, sever their human relationship with targets of praise and blame (at least, while they are praising or blaming them).

Notably, even Vargas' moderately revisionist account accommodates for radical instrumentalists. Vargas admits that it is possible for one to adopt an "external standpoint" relative to the practice if, for instance, they doubt whether a particular application of a backward-looking norm is truly justified.³⁶ In these moments, forward-looking considerations come to the forefront of one's deliberation over an application of the backward-looking norms of responsibility. For instance, we might ask whether praising or blaming some person in a particular situation, will yield certain effects. However, adopting this external standpoint need not be prompted by doubt. It is completely compatible with Vargas' account that one be a radical instrumentalist and apply the backward-looking norms with only forward-looking considerations in mind. For example, a radical instrumentalist might blame their friend for stealing their wallet in accordance with backward-looking norms but be motivated only by the prospects of cultivating more moral considerations-responsive agents. Further, that they are motivated to blame with only forward-looking considerations in mind does not compromise the potential for those instances of praise and blame to achieve the effects that, for Vargas, justify the practice. It seems that radical instrumentalists can bring about the justifying effects of the practice just as well, if not better, than standard issuers of praise and blame. Thus, on Vargas' account, radical instrumentalists may be justified in praising and blaming as they do. It is justified, on Vargas' account, to hold the objective attitude towards targets of praise and blame. I argue that this is not the case on the game view.

³⁶ Vargas, *Building Better Beings*, 184.

Notably absent from radical instrumentalists of any game is the lusory attitude. Holding the lusory attitude towards the rules of some rule-guided activity is necessary for genuine gameplay – a genuine gameplayer must adhere to the rules *because* the rules make the game possible. Radical instrumentalists are not playing the responsibility game, because they are not adhering to the rules for the sake of the possibility of gameplay; they are not accepting the backward-looking norms because they want to be able to praise or blame. Radical instrumentalists may accept the backward-looking rules of the responsibility game only insofar as they are conducive to some other end. Now, being a radical instrumentalist is not necessarily wrong, all things considered. It is just the case that the radical instrumentalist would be incorrect in insisting that they are praising or blaming. Instead, they are performing some out-of-game action (such as moral education or, more maliciously, manipulation) whose justification rests on out-of-game normative considerations. To take a previous example, it may well be a good thing, all things considered, that the radical instrumentalist ran the 100m track only to defuse the bomb just as it may well be an all-things-considered good thing that the radical instrumentalist blames their friend only for some greater good. An instance of radical instrumentalist “blame” may be justified in light of out-of-game normative considerations, but it is not an instance of blame and so is not an instance of justified blame. The act occurs out-of-game, leaving the normative background of the responsibility game. Genuine gameplayers must exhibit the lusory attitude towards the norms that guide the practice – they must praise and blame for the sake of praising and blaming. Thus, unlike Vargas’ and more radically revisionist accounts, the game view has the benefit of providing conceptual resources to distinguish between radical instrumentalists and genuine gameplayers by its requirement of the lusory attitude. By the same requirement, the game view also ensures that genuine practitioners (or gameplayers) of responsibility preserve the

interpersonal human relationships with targets of praise or blame, sparing these targets from the objective attitude.

One might object that the gameplayer-instrumentalist distinction and the game view's requirement of the lusory attitude are only valuable if we assume that a responsibility practice guided by backward-looking norms is more valuable than a forward-looking practice. If I am assuming that Strawson's critique of the objective attitude extends to forward-looking norms just in light of their being forward-looking, then it seems as though I am begging the question against radically forward-looking theorists. Otherwise, without the assumption, the distinction between radical instrumentalists and genuine gameplayers seems superfluous. I must admit that I do assume that it is preferable to be held responsible in such a way that the person holding us responsible does not hold the objective attitude towards us. However, this assumption only holds that *being motivated* to engage with our responsibility norms in a certain way (by, say, holding the lusory attitude towards them) is more valuable than a different way of engaging with them. I am not assuming that a certain set of responsibility norms are more valuable than another set of responsibility norms. One does not exhibit the objective attitude in virtue of following a certain set of responsibility norms. Rather, one exhibits the objective attitude when they engage with (any set of) responsibility norms in a certain way. Thus, the distinction is beneficial insofar as it allows us to differentiate between ways of engaging with our responsibility norms rather than between sets of responsibility norms that differ in value. Further, the requirement of the lusory attitude is beneficial insofar as it ensures that those who participate in our responsibility practices engage with the responsibility norms in a certain way rather than engage with a certain set of responsibility norms. This objection is misguided.

5.2. The Genuine Gameplayer

The previous sub-section discusses a conceptually useful distinction available to the game view. This distinction between non-gameplayers and genuine gameplayers ensures that targets of praise and blame are spared from the objective attitude (else they risk being a radical instrumentalist). There is yet another benefit to playing the responsibility game – a benefit that comes from being on the right side of a distinction between the non-gameplayer and the genuine gameplayer. Consider a complaint against consequentialist moral theories about their “unlimited scope”, as Onora O’Neill puts it.³⁷ That is, because such theories prescribe the maximization of some property (i.e. happiness) from an action’s consequences, identifying morally right actions requires that we have knowledge of the consequences of an arguably infinite number of possible actions. Even if we do bring about good consequences through our actions, there seems to be an ever-lingering possibility that a different action could have brought about better consequences. The application of such theories, in this regard, resembles a type of activity that Suits takes to be the opposite of a game: a “technical activity”.³⁸ Technical activities are, like games, goal-directed activities. Unlike games, however, they are done solely to achieve the goal towards which they are directed. An oft-cited example by Suits of a technical activity is “work”, broadly speaking, or a job done solely to achieve material gain – such as money, for instance. If one could afford to slack off during work or had the magical ability to time-skip to 5pm, one certainly should take the opportunity to employ these more efficient means of securing material gain. If, while subscribing to a consequentialist moral theory, one knew how to act in a way that would bring about better consequences than another act, one certainly should perform the better act. If one knew of better means to achieve the goal of a technical activity (perhaps through, say,

³⁷ O’Neill, “A Simplified Account of Kant’s Ethics,” in *Contemporary Moral Problems*, 4th edition, ed. James White (St. Paul: West Publishing Company, 1994), 46.

³⁸ Suits, *The Grasshopper*, 22.

occultist books on time control), one would be “technically intelligent” to employ such means over less efficient means.³⁹ It is constitutive of technical activities that the means through which one pursues their goals are justified in proportion to their efficacy at achieving said goals. Success within a technical activity is restricted by one’s knowledge about the arguably infinite possible actions that can bring about the goals of the technical activity. Accordingly, for every action taken within a technical activity, there seems to be the ever-lingering possibility that there is a different, better action that could have been taken and – in virtue of being better – an action that should have been taken. Let us refer to this feature of technical activities as “efficacy anxiety”.

The tight justificatory connection suggested by Vargas and radical forward-looking theorists between the norms and the goal of our responsibility practices reveal that both camps hold what I will call a technical activity view of moral responsibility. Both types of theorists agree that the means used to achieve the goal of our responsibility practices are justified in proportion to their efficacy at doing so. Dispute arises between these two theorists largely over which set of practices is more efficacious; Vargas insists that a practice guided by backward-looking norms would be more efficacious than a solely forward-looking one, while Pereboom suggests otherwise. Applying either theory, however, will exhibit an efficacy anxiety like that of consequentialist moral theories. For every instance of praise or blame on these accounts, there is a lingering possibility that there is another, better way that one could and should have praised or blamed.

Unlike technical activities, games do not prescribe participants to take the most efficacious actions towards the prelusory goal (or risk being “technically unintelligent”). Recall

³⁹ Ibid., 23.

that it is even constitutive of games to prohibit more efficient means in favor of less efficient means of pursuing their prelusory goal. In a game, all we are “supposed to do” – all that is required of us – is to play. Actions within games are not justified in proportion to their efficacy at achieving the goal of the game. Actions are justified just as long as the player is genuinely playing the game – striving towards the prelusory goal while adhering to the rules for the sake of playing. Unlike our responsibility practices under a technical activity view, there is no lingering possibility of some other, better way that one could and should play the game. Thus, efficacy anxiety finds no place within games, including the responsibility game.

The absence of efficacy anxiety illuminates another benefit of the game view. In conjunction with other features of games, the absence of efficacy anxiety allows one to “wholeheartedly” participate in a game in a way they cannot participate in a technical activity.⁴⁰ William Morgan similarly points out that “games make it possible for players to utterly absorb themselves in the actions they take to realize game goals”.⁴¹ I propose that games do this by simplifying agency and promoting a sense of ownership over in-game actions. First, games direct its players towards its prelusory goal and restrict the means through which its goal can be pursued. This first restrictive layer describes the choices that genuine gameplayers can take in playing the game, simplifying one’s agency by placing constraints on the possible ways to exercise it. Thi Nguyen writes on this feature of games, “One of the greatest pleasures games offer is a certain existential balm—a momentary shelter from the existential complexities of ordinary life. In a game, for once in my life, I know exactly what it is that I’m supposed to be doing”.⁴² Admittedly, games can be more or less complex. In chess, especially for the novice, one might not “know exactly

⁴⁰ William Morgan, “Gameplay, wholehearted engagement, and the good life,” *Journal of the Philosophy of Sport* 48, no. 3 (2021): 366.

⁴¹ *Ibid.*, 356.

⁴² C. Thi Nguyen, *Games: Agency as Art*, (New York: Oxford University Press, 2020), 67.

what it is that one is supposed to be doing” to achieve a checkmate, even if they know how the pieces move. Similarly, a responsibility game may be rather complex, even within the constraints of quality-of-will rules; one might not know exactly how to best build better beings. However, in conjunction with the absence of efficacy anxiety, we can recognize that the complexity of *winning* does not necessarily make *playing* complex. Again, unlike technical activities, games only prescribe that we *try* to achieve a state of affairs. While one may not know how to achieve a checkmate, one at least knows that they have to try to achieve a checkmate and the valid moves they can take to get there. One need not deliberate over an overwhelming array of actions to choose what will be most efficient at arranging the chess board in a certain way (e.g. restraining my opponent’s hands so I can do so without resistance, paying the officials to add time to my timer, etc.). Unlike technical activities, games simplify our agency by constraining the possible ways to exercise our agency. Our agency is further simplified by the absence of efficacy anxiety, dampening the demands for how our agency should be exercised.

Second, games also restrict the ways in which gameplayers engage with the rules that define the means of pursuing the game’s goals; gameplayers must exhibit the lusory attitude (else players risk being radical instrumentalists). This second restrictive layer gives each player a sense of ownership over their actions. Admittedly, this claim may seem at odds with the restrictions that games place on valid actions, for it seems like how free a choice is should be positively correlated with the amount of choices available to the chooser. However, as Nguyen also observes, this concern presupposes a “negative view of autonomy, in which we support the autonomy of others by leaving them alone”.⁴³ I do not seek to propose an objection to the negative view of autonomy nor defend an alternative view, but I can settle for admitting that

⁴³ Nguyen, *Games*, 77.

those who are at least sympathetic with a more positive sense of autonomy will recognize that games do not preclude autonomy in this sense – one that holds that people are acting voluntarily when they act how they want to. What the requirement of the lusory attitude allows us to appreciate is that, as Gingerich writes, “if I am engaged in an activity that is recognizable as ‘playing a game,’ I am engaged in an activity that I do not have to do, and that I regard as such”.⁴⁴ Because the possibility of a game must be among the player’s reasons for playing it, playing a game can only happen when a putative player wants to play it. The absence of efficacy anxiety further deepens the player’s sense of ownership over their in-game actions. Because games only prescribe that we play, what we do within games cannot be settled by what Julia Annas terms a “decision procedure, a systematic and theorizable way of telling us what to do” (although, there may be one for winning a game).⁴⁵ At most, game rules tell us what we *can* do, but our in-game actions are, for the most part, left undetermined by game rules. Unlike actions within technical activities, in-game actions are largely settled by the player rather than dictated by what is most efficient.

The foregoing discussion reveals that the absence of efficacy anxiety on the game view allows us to appreciate a way of blaming and praising wholeheartedly in the sense that it both simplifies participants’ agency and ensures participants a sense of ownership over how they practice their agency in-game. On the game view, praising and blaming need not be overwhelmingly complex. We need not consider the countless ways of building better beings to settle on the best one. Further, our praise and blame can be something we can feel ownership over. Whether one praises and blames and how one praises and blames is, ultimately, up to them.

⁴⁴ Jonathan Gingerich, “Freedom and the Value of Games,” *Canadian Journal of Philosophy* 48, no. 6 (December 1, 2018): 9.

⁴⁵ Julia Annas, “Being Virtuous and Doing the Right Thing,” *Proceedings and Addresses of the American Philosophical Association* 78, no. 2 (2004): 63.

This wholehearted pursuit of a goal that we engage in when we play games, according to Morgan, is an “important mark of the good life”.⁴⁶ Thus, the game view also allows us to appreciate that praising and blaming can contribute to our pursuit of the good life by allowing us to participate in the wholehearted pursuit of a goal – or, at the very least, a pursuit more wholehearted than that afforded by a technical activity view.

Importantly, unlike technical activity views, the benefits of the game view do not find their source of value from the achievement of the aim or goal of the game. This is a common feature of games, which usually have rather trivial aims – e.g. reaching a finish line before other participants, putting a ball into a basket more times than another team, putting your opponent in checkmate. Games derive their value from the *process* of trying to achieve their aim. Thomas Hurka notes that, in this sense, games subvert the “classical” or Aristotelian ideas of value.⁴⁷ He argues that games cannot be classified under either of the two Aristotelian models of value for activity: *kinesis* and *energeia*. According to Hurka, a *kinesis* “is an activity aimed at a goal external to it” while an *energeia* “is not directed at an external goal but has its end internal to it”.⁴⁸ He takes the activity of contemplation as an example of an *energeia* “because it does not aim to produce anything beyond itself”.⁴⁹ In light of these differences, Aristotle, on Hurka’s reading, “assumed that the value of a *kinesis* must derive from that of its goal”. *Energeia*, on the other hand, derives its value from mere participation in it. Thus, these two types of activity can be distinguished by their form and their “value base”. Formally, if *x* is a *kinesis*, then *x*-ing will lead one to either successfully *x* or otherwise fail to *x*. This is not true of an *energeia*.

⁴⁶ Morgan, “Gameplay, wholehearted engagement, and the good life,” 356.

⁴⁷ Thomas Hurka and John Tasioulas, “Games and the Good,” *Proceedings of the Aristotelian Society* (Hardback) 106, no. 1 (June 2006): 229.

⁴⁸ Ibid.

⁴⁹ Ibid.

Contemplating, for example, will necessarily lead one to successfully have contemplated. Further, because a kinesis is aimed at bringing some state of affairs about, its value base is located in the achievement of bringing this state of affairs about; a kinesis is valuable when participation in it realizes its aims and in proportion to the value of its aims. Conversely, the value base of an *energeia* is located within the *energeia* itself; mere participation in the *energeia* is valuable.

Games subvert this classical distinction by having the form of a kinesis while having the value base of an *energeia*. Consider again the 100m dash. Trying to cross the finish line before other participants will lead one to successfully do so or fail to do so. However, despite having the form of a kinesis, we would deny that the value of the 100m dash lies within the achievement of its prelusory goal. The 100m dash is not valuable only insofar as one actually crosses the finish line before other participants. Rather, there are features of the process of achieving the prelusory goal that render mere participation in it valuable. It may be, as Hurka argues, that the difficulty of trying to achieve the prelusory goal renders it valuable. Perhaps, as John Tasioulas suggests, participation may be rendered valuable by the social dimension of play – “that there are spectators or competitors who can or do ‘play along’, engaging in or valuing the activity as worthwhile”.⁵⁰ Or, lastly, as Jonathan Gingerich argues, the freedom that one experiences in the non-compulsory pursuit of a goal might render such participation valuable.⁵¹

Consider now the responsibility game. Building better beings will possibly lead one to successfully build better beings and possibly fail to build better beings. However, unlike a kinesis, I have argued that there is value to be located within the game itself – mere participation

⁵⁰ Ibid., 247

⁵¹ Gingerich, “Freedom and the Value of Games,” *Canadian Journal of Philosophy* 48, no. 6 (December 1, 2018), 8-9.

in the process of building better beings is valuable. In this section, I argued that there are reasons why mere participation in the responsibility game is valuable over and above a responsibility practice under a technical activity view. Such a practice preserves our interpersonal human relationships with targets of praise and blame, and enables and ensures wholeheartedness in our praising and blaming. To be clear, I do not deny that building better beings has value; I only deny that building better beings is the only reason we have to praise and blame. So, unlike many games with rather trivial aims, the responsibility game has a rather valuable aim – one worth achieving on its own. But what the game view allows us to appreciate is that there is *also* value within the process of achieving it. Building better beings is certainly a good reason to play the game of praise and blame just as winning a bunch of money is a good reason to play poker. However, I hope the game view has shown that playing *itself* gives us good reasons to do so as well. To use Hurka’s words, the game view shifts the focus towards “process rather than product, journey rather than destination”.⁵²

⁵² Hurka and Tasioulas, “Games and the Good,” 217.

6 A RESPONSE TO AN OBJECTION AND A QUALIFICATION

At this point, it is worth addressing an objection that disfavors any practice of responsibility guided by backward-looking norms, such as the one suggested in the game view described above and the one Vargas insists on in his account. The heart of the objection is that we seem to have good reason not to participate in any responsibility practice guided by backward-looking norms, reasons independent from considerations of efficacy. Pereboom cites empirical evidence that the moral anger characteristic of – and often justified by – the backward-looking norms of moral responsibility leads to destructive effects and argues that these effects can be avoided if we adopt a practice guided by forward-looking norms.⁵³ Under a forward-looking account, moral anger may also hinder the effects that justify the practice, rendering a certain degree of moral anger unjustified. Thus, forward-looking practices can regulate moral anger and avoid its other dysfunctional downstream effects. Pereboom concludes, in light of considerations besides efficacy, we have more reason to engage in a forward-looking practice rather than a backward-looking practice. In this section, I will respond to this objection and then discuss the limits of what I have argued for so far.

6.1. A Response to an Objection

In response, I suggest that those who express excessive moral anger are, like radical instrumentalists, not genuine gameplayers and thus leave the normative background of the game. Suits describes another class of pseudo-gameplayers, the trifler, who may follow the rules of a game but neglect the game's prelusory goal. A trifler at chess, for example, may "simply be trying to get six of his pieces to the other side of the board before he is checkmated" or "may just

⁵³ Although Pereboom does discuss the destructive effects of anger that are relevant to the efficacy of attaining the prelusory goal of the responsibility game, the point is that he also cites reasons independent of these concerns. To name a few, he notes that anger impairs relationships, "enhances the disposition to view ambiguous behavior as hostile", and "strengthens the proclivity to rely on stereotypes about irrelevant features such as race and ethnicity".

be moving his pieces at random”.⁵⁴ The trifler at chess acts according to the rules of chess but does not adopt the prelusory goal of putting their opponent in checkmate. Like being a radical instrumentalist, being a trifler is not necessarily all-things-considered wrong. However, triflers would be incorrect in insisting that they are playing the game. Their actions occur out-of-game and, thus, their justification cannot appeal to in-game rules. For instance, the trifler at chess may admit to their opponent that they are moving their pieces completely randomly, and their opponent may rightfully wonder why the trifler does this. It would be an unsatisfactory response if the trifler were to simply appeal to the fact that the rules of chess allow them to act as they do. The trifler may, however, appeal to other considerations – perhaps the trifler was tired of playing chess or thought the game was a waste of time. Either way, appealing to the rules of chess is not sufficient to justify the trifler’s actions because they are not really playing chess; they leave the normative background of the game rules and now must appeal to broader normative considerations in justifying their actions.

I claim that those who express destructive moral anger during instances of blame are triflers with respect to the responsibility game. In exhibiting destructive moral anger, these triflers are acting within the rules with the purpose of expressing their anger, asserting their control of the target of their blame, or in the service of some other end rather than a means for building better beings. The triflers, however, do not adopt the prelusory goal of the responsibility game. This illuminates the sense in which moral anger is regulable without forward-looking norms. Much like the chess trifler, the actions of the responsibility game trifler occur out-of-game and cannot appeal to the rules of the responsibility game for justification. They must appeal to broader normative considerations. However, given the destructive effects of unregulated moral anger that

⁵⁴ Suits, *The Grasshopper*, 46.

Pereboom mentions, it seems there are broader normative considerations that count against unregulated moral anger. Thus, the trifler does have reason to regulate their moral anger even if the responsibility game is guided by backward-looking rules – the trifler must answer to out-of-game normative considerations.

6.2. A Qualification

Now, I have argued that we have reasons to play the game of praise and blame and have argued against a reason we might have to not play it. However, it is worth acknowledging the limits of what I have argued for. The benefits discussed (that we avoid the objective attitude and praise and blame wholeheartedly on the game view) are benefits only over and above a technical activity view. This means that other putative responsibility games, ones guided by forward-looking rather than backward-looking rules or ones with a different prelusory goal (i.e. to proportionately harm those who do harm), can claim these same benefits over and above a responsibility practice that is a technical activity. Here, I do not address the possibility of other plausible responsibility games, nor do I address the benefits that the responsibility game fleshed out here might have over and above other responsibility games. Embarking on such a cost-benefit analysis between different responsibility games is not within the scope of this paper. However, I suggest that, given that the above responsibility game can fulfill the desiderata for a middle path between backward-looking and forward-looking notions of responsibility and the value of fostering moral considerations-responsive agency, a responsibility game with quality-of-will rules and the prelusory goal of building better beings is a particularly attractive candidate for our participation. However, I happily leave open this question for future research.

7 CONCLUDING REMARKS

In this paper, I proposed an alternative to Vargas' theory of responsibility that can fulfill the three desiderata of a "middle path" between backward-looking and forward-looking notions of responsibility. The game view, which proposes that our responsibility practices are a game, can preserve the ordinary backward-looking responsibility norms if they constituted the rules of the responsibility game. At the same time, the game view can avoid the objection from inefficacy, because if our responsibility practices are a game then participation in its practices is not necessarily justified in proportion to its efficacy at achieving its goal. Further, I argued that we have good reasons to play the game, thus, the game view can justify our responsibility practices without reference to a controversial control condition. I argued that the game view preserves our interpersonal human relationships with targets of praise and blame. Then, I argued that, due to the absence of efficacy anxiety, the game view also allows praise/blame to constitute the wholehearted pursuit of a goal which leads us closer to the good life. Further, I argued against a potential reason to not play the game: Pereboom's concern that backward-looking responsibility practices cannot regulate destructive moral anger. I argued that destructive moral anger is not really playing the responsibility game and, thus, should be regulated in light of out-of-game normative considerations. Ultimately, the game view allows us to appreciate that both the aim and the *means* we take towards the aim of our responsibility practices give us good reason to praise and blame.

8 BIBLIOGRAPHY

- Annas, Julia. "Being Virtuous and Doing the Right Thing." *Proceedings and Addresses of the American Philosophical Association* 78, no. 2 (2004): 61–75.
<https://doi.org/10.2307/3219725>.
- Bhanji, Jamil P., and Mauricio R. Delgado. "The Social Brain and Reward: Social Information Processing in the Human Striatum." *Wiley Interdisciplinary Reviews: Cognitive Science* 5, no. 1 (October 8, 2013): 61–73. <https://doi.org/10.1002/wcs.1266>.
- Caruso, Gregg, and Derk Pereboom. *Moral Responsibility Reconsidered*. Cambridge University Press, 2022.
- Coren, Daniel. "Moral Responsibility Must Look Back." *American Philosophical Quarterly*, n.d.
- Doris, John M. "Doing without (Arguing About) Desert." *Philosophical Studies* 172, no. 10 (April 11, 2015): 2625–34. <https://doi.org/10.1007/s11098-015-0477-2>.
- Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. *Four Views on Free Will*. Malden (Ma) ; Oxford ; Victoria: Blackwell, 2014.
- Gingerich, Jonathan. "Freedom and the Value of Games." *Canadian Journal of Philosophy* 48, no. 6 (December 1, 2018): 831–49. <https://doi.org/10.1080/00455091.2017.1423224>.
- Hurka, Thomas, and John Tasioulas. "Games and the Good." *Proceedings of the Aristotelian Society (Hardback)* 106, no. 1 (June 2006): 217–35. <https://doi.org/10.1111/j.0066-7373.2006.00132.x>.
- Jean-Richard-Dit-Bressel, Philip, Simon Killcross, and Gavan P. McNally. "Behavioral and Neurobiological Mechanisms of Punishment: Implications for Psychiatric Disorders." *Neuropsychopharmacology* 43, no. 8 (July 1, 2018): 1639–50.
<https://doi.org/10.1038/s41386-018-0047-3>.

- McCormick, Kelly Anne. "Anchoring a Revisionist Account of Moral Responsibility." *Journal of Ethics and Social Philosophy* 7, no. 3 (June 5, 2017): 1–20.
<https://doi.org/10.26556/jesp.v7i3.76>.
- McGeer, Victoria. "Building a Better Theory of Responsibility." *Philosophical Studies* 172, no. 10 (April 14, 2015): 2635–49. <https://doi.org/10.1007/s11098-015-0478-1>.
- Morgan, William J. "Game Play, Wholehearted Engagement, and the Good Life." *Journal of the Philosophy of Sport* 48, no. 3 (June 30, 2021): 1–13.
<https://doi.org/10.1080/00948705.2021.1948337>.
- Nguyen, C. Thi. *Games: Agency as Art*. New York: Oxford University Press, 2020.
- O'Neill, Onora. "A Simplified Account of Kant's Ethics." In *Contemporary Moral Problems*, edited by James White. St. Paul: West Publishing Company, 1994.
- Pereboom, Derk. "Undivided Forward-Looking Moral Responsibility." *The Monist* 104, no. 4 (September 4, 2021): 484–97. <https://doi.org/10.1093/monist/onab014>.
- Shoemaker, David, and Manuel Vargas. "Moral Torch Fishing: A Signaling Theory of Blame." *Noûs* 55, no. 3 (October 24, 2019): 581–602. <https://doi.org/10.1111/nous.12316>.
- Smart, J. J. C. "Free Will, Praise and Blame." *Mind* 70, no. 279 (1961): 291–306.
<https://doi.org/10.1093/mind/lxx.279.291>.
- Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75, no. 1-2 (August 1994): 5–24. <https://doi.org/10.1007/bf00989879>.
- Strawson, P. F. "Freedom and Resentment." In *Agency and Responsibility: Essays on the Metaphysics of Freedom*, edited by Laura Waddell Ekstrom. Boulder: Westview, 2000.
- Suits, Bernard. *The Grasshopper: Games, Life and Utopia*. Toronto: University of Toronto Press, 1978.

———. “What Is a Game?” *Philosophy of Science* 34, no. 2 (June 1967): 148–56.

<https://doi.org/10.1086/288138>.

Vargas, Manuel. *Building Better Beings: A Theory of Moral Responsibility*. New York: Oxford University Press, 2013.

Vargas, Manuel . “Desert, Responsibility, and Justification: A Reply to Doris, McGeer, and Robinson.” *Philosophical Studies* 172, no. 10 (April 26, 2015): 2659–78.

<https://doi.org/10.1007/s11098-015-0480-7>.

———. “Moral Influence, Moral Responsibility.” In *Essays on Free Will and Moral Responsibility*, edited by Nick Trakakis and Daniel Cohen, 91–123. Newcastle: Cambridge Scholars Press, 2008.