

## Responsibility and False Beliefs

Peter Vallentyne, University of Missouri-Columbia

in *Justice and Responsibility*, edited by Carl Knight and Zofia Stemploska (Oxford University Press, 2011), pp. 174-86.

An individual is agent-responsible for an outcome just in case it flows from her autonomous agency in the right kind of way. The topic of agent-responsibility is important because most people believe that agents should be held morally accountable (e.g., liable to punishment or having an obligation to compensate victims) for outcomes for which they are agent-responsible and because many other people (e.g., brute luck egalitarians) hold that agents should *not* be held accountable for outcomes for which they are *not* responsible. In this paper, I examine how false beliefs affect agent-responsibility.

Unlike most of the papers in this collection, my paper is on the notion of agent-responsibility that many believe is relevant to justice and morality generally. I do not here address the question of how, if at all, justice and morality are sensitive to agent-responsibility.

### 1. The Concept of Agent-Responsibility

The term “responsibility” is used in various ways and *agent-responsibility* (or attributive responsibility) needs to be clearly distinguished from the others. The rough idea is that an individual is agent-responsible for a choice or outcome to the extent that it suitably reflects the exercise of her autonomous agency. Being *causally responsible* for an outcome, for example, is necessary but not sufficient for agent-responsibility, since an agent may reasonably have been

unaware that her choice had the effect in question (e.g., when one flicks a light switch that has been secretly rigged to set off a bomb).

Moreover, an agent's being agent-responsible for a choice or outcome does not entail that she is *morally accountable* (or substantively responsible) for the choice or outcome in the sense that she has certain substantive moral duties, liabilities, etc. in virtue of the occurrence of the outcome (e.g., has the duty to compensate those harmed by her actions, or is subject to permissible punishment). That is a substantive moral question that will be answered in different ways by different moral theories. Agent-responsibility merely establishes that the choice or outcome flows in the right kind of way from the individual's agency. It leaves open what the substantive moral implications are. Although most substantive moral theories hold agents morally accountable only for that for which they are agent-responsible, some do not.<sup>1</sup>

An agent is agent-responsible for some outcome to the extent that it suitably reflects her exercise of agency. Such an outcome is part of the agent's record (or ledger) concerning her exercise of autonomous agency and thus relevant to the credit-worthiness of that exercise.<sup>2</sup> Reactive attitudes (blame, praise, etc.) towards an agent with respect to an outcome are appropriate in principle if and only if the agent is agent-responsible for the outcome.<sup>3</sup> Under normal circumstances, for example, it is inappropriate to blame a person for the fact that it rained.

The mere fact that an individual is agent-responsible for an outcome does not establish that any particular reactive attitude is appropriate. First, the appropriate reactive attitude will depend on the normative perspective at issue (e.g., morality vs. prudence). An agent might be prudentially praiseworthy for stealing the money (with little chance of being caught) while being morally blameworthy. Second, even for a given normative perspective (e.g., morality), the issue

of what determines blameworthiness and praiseworthiness is a substantive issue. That an agent is agent-responsible for an outcome merely establishes that it is not inappropriate in principle to have reactive attitudes towards the agent with respect to the outcome. It does not establish what attitudes are appropriate.<sup>4</sup>

Although it is not standard, we need, I believe, to distinguish between a narrow and broad concept of agent-responsibility. Consider two almost identically situated agents, each of whom throws a rock at my window with the intention of breaking it. Both believe that it is certain that her rock will break the window and this belief is true for one agent but false for the other. In fact, for the second agent, it is certain that the rock will not break the window (e.g., because, unbeknownst to her, the rock is too fragile). For what outcomes is each agent responsible?

If we are concerned with assessing—in the strictest sense—the merits (e.g., appropriate moral or prudential reactive attitudes) of each individual’s exercise of agency, then neither agent, I would argue, is fully responsible for breaking the window. One agent did not break my window at all. The other agent did break my window, but that event depended in part on factors that do not reflect her exercise of agency (e.g., that her rock was hard enough to break my window). That difference is not a reflection of the agency of either individual. Moreover, the normatively (e.g., morally or prudentially) appropriate reactive attitudes (blame, praise, etc.), I would argue, are not sensitive to the presence or absence of factors that the agent could not reasonably have known about. In the strictest sense, I believe, agents are responsible only for the believed (intended, etc.) outcomes of their choices.<sup>5</sup> Call this *narrow responsibility*. In the example, the two agents have the same narrow responsibility. It is, of course, controversial that, in the strictest sense, the appropriateness of reactive attitudes (and assessment of desert) depends solely on narrow responsibility. I shall not attempt to defend this claim here. The important point to note is

that, for narrow responsibility, the truth or falsity of a belief is irrelevant. Agents are responsible, at most, for the believed outcomes of their choices—whether or not they are true.<sup>6</sup> They are never morally narrowly responsible for objective outcomes.

Narrow responsibility is not, however, the topic that I wish to address. I am interested in the notion of responsibility that connects the exercise of autonomous agency with objective outcomes in the world. Call this *broad responsibility*. It is intermediate between narrow responsibility (based solely on the agent's mental states) and causal responsibility (based solely on the outcomes of choices). Roughly speaking, agents are broadly responsible for an outcome to the extent that they are causally responsible for it and it suitably reflects the exercise of their autonomous agency. Like casual responsibility, it is concerned with responsibility for events in the external world (as well as mental states). Like narrow responsibility, it holds that an agent's beliefs about the outcomes of her choices may limit that for which she is responsible.<sup>7</sup> In the above example, both agents made an autonomous choice to throw a rock at my window with the intention of breaking it. The first agent broke my window, foresaw this result, and is thus, we may suppose, broadly responsible for the breaking of my window. The second agent did not break my window and thus is not broadly responsible for its breaking.

Broad responsibility is the notion that most have in mind when they claim that individuals are morally accountable for those outcomes for which they are responsible.<sup>8</sup> In what follows, I focus on broad responsibility and all unqualified references to responsibility or agent-responsibility should be so understood.

In order for an individual to be (broadly) responsible for an outcome, three general conditions must be satisfied: (1) an autonomy condition: she must make a suitably autonomous choice, (2) a causal condition: the outcome must be suitably causally related to the choice, and

(3) a belief condition: she must have (or should reasonably have) a suitable belief that the outcome is so related. I shall assume throughout that the autonomy condition is satisfied. My focus below will be on the third condition. First, however, I shall briefly comment on the causal condition.

I shall assume that the causal condition restricts responsibility, relative to a choice, to the impact the choice has relative to some background baseline for the choice situation. The key point is that the agent (in the simple cases we address) is not responsible for the baseline choice situation. She is only responsible for the difference her choice makes relative to the baseline for the choice situation. Call this—the difference between the baseline and the outcome of her choice—the *impact* of her choice. I leave open how that baseline is determined, except to stipulate that it is suitably sensitive to the agent's disposition, given her beliefs and choice-making abilities, to choose and the resulting consequences. As a very crude approximation, we might think of the baseline as that which would happen if the agent made a reasonable choice (given her beliefs and choice-making abilities) in the choice situation. Thus, for example, suppose that only money matters, that one agent faces a choice between \$10 and \$100, that another faces a choice between \$10 and \$20, and that both choose the \$10. Suppose that the baseline for the first agent is \$90 (90% of best) and that for the second agent is \$18 (90% of the best). The impact of choosing the \$10 for the first agent would be -\$80, whereas the impact of choosing the \$10 for the second agent would be -\$8.

Of course, agents are typically partly responsible for their choice situations (their beliefs, desires, or opportunities) in virtue of prior choices. When this is so, the agents are also partly responsible for their baseline (as well as for the impact their choices make relative to the baselines). A full account of responsibility for outcome will be historical and allocate

responsibility for a given outcome on the basis of *all* the autonomous choices that an agent has made. Here, however, I shall address only the simple case where an agent is not responsible for her baseline (e.g., because it is her first choice). In particular, I assume that agents are not responsible for their false beliefs.<sup>9</sup> In this simple case, agents are responsible, I assume, only for the impact of their choices, that is, the differences their choices make relative to their choice-context-sensitive baselines.

## 2. False Beliefs

How do false beliefs affect responsibility? They do not affect *causal* responsibility, since the agent's *beliefs* are irrelevant to that topic. Nor do they affect *narrow* agent-responsibility, since the *truth* of the agent's beliefs is irrelevant to that topic. For broad agent-responsibility, however, false beliefs can affect responsibility. For example, an agent is not broadly agent-responsible for an outcome if there was no way she could have known that her choice would produce the outcome. In what follows, I shall explore how false beliefs affect (broad agent-) responsibility.

False beliefs, I claim, can affect responsibility in three distinct ways. One is that false beliefs affect an agent's disposition to choose and can thereby affect the baseline for the allocation of responsibility. The baseline for an agent who falsely believes (with no responsibility for the belief) that smoking is good for health is different (e.g., lower) than that for an identical agent but with no false beliefs (because the former is more likely to smoke). This is a very important topic, but an account of how the baseline is determined (and affected by false beliefs) is beyond the scope of this paper.<sup>10</sup>

A second way that false beliefs affect responsibility is by inducing *unforeseen outcomes* (relative to the baseline). This is that part of the outcome that the agent did not foresee. Suppose,

for example, that an agent, under normal conditions, flicks a light switch and thereby, unbeknownst to her, sets off a bomb. Setting off the bomb is an unforeseen outcome of her choice. It's relatively uncontroversial that, where agents do not foresee an outcome and are not responsible for the failure to foresee it, they are not responsible for that outcome. Such unforeseen outcomes play no role in choice deliberations and agents are thus not responsible for them. I have nothing to add to this case.

A third way that false beliefs affect responsibility is by inducing *imaginary outcomes*. This is that part of what the agent believes the outcome of her choice to be that was not part of the causal outcome. For example, if an agent flicks the lights switch believing that it would cause the room to light up, but in fact the switch is broken, then the lighting of the room is an imaginary outcome of her choice. Imaginary outcomes are outcomes anticipated by the agent that are not real, whereas unforeseen outcomes are real outcomes that the agent did not anticipate. In this chapter, I focus on how the existence of imaginary outcomes affects responsibility for outcomes.

It's important to distinguish imaginary outcomes from unrealized foreseen objective possibilities. When an agent makes a choice that she accurately foresees will raise the objective chance of winning the lottery, but she does not win, the possibility of winning is not an imaginary outcome. It is a real possibility. Whether such a possibility is realized is a matter of option luck—roughly: luck in how foreseeable and influenceable objective chances are realized. Agents who have complete and true beliefs are still subject to option luck in a world with objective chances. I address this topic elsewhere.<sup>11</sup> Here our topic concerns false beliefs, not unrealized foreseen objective chances. Imaginary outcomes are believed outcomes with no objective basis. They can arise independently of objective chances. For simplicity, I shall assume

throughout that the outcomes of choices are objectively fully determined (with no objective chances other than 0 or 1), but similar points apply for cases where the outcomes are not so determined.

Our question concerns how imaginary outcomes affect responsibility for outcomes.

### 3. Imaginary Outcomes

How is the imaginary outcome of a choice relevant for the determination of responsibility?

Because the imaginary outcome is not part of the actual outcome, one might suppose that it is irrelevant to broad responsibility. I shall suggest that this is sometimes, but not always, so.

To see that the imaginary outcome is sometimes relevant to responsibility, consider a case in which a benevolent doctor performs an operation that she knows will be extremely painful. She performs the operation because she believes that this will save the patient's life and ultimately be good for the patient. In fact, the patient will die no matter what the doctor does, and thus the net effect of the doctor's action is to increase very significantly the pain experienced by the patient. Is the doctor responsible for the increase in pain? On the one hand, it seems that she is, since she imposed it intentionally. On the other, she is at least partly excused, since she did it falsely believing that it would benefit the patient. With respect to moral blame, she is certainly less blameworthy than she would be if she had done it without the imaginary benefit. Thus, the imaginary outcome is, at least sometimes, relevant to responsibility. At a minimum, it can sometimes mitigate responsibility for outcomes. It's not clear, however, how it does this.

In the case of unforeseen outcomes, we simply strip them away from the full causal outcome and "record" only the foreseen impact in the responsibility account. A comparable move is not possible for imaginary outcomes. Imaginary outcomes are not real outcomes and



hence cannot be stripped away. Nor can they be simply added to the foreseen outcome. That would give us the believed outcome, which is relevant for narrow responsibility but not for broad responsibility. Nor can some of the foreseen outcome be stripped away in virtue of the imaginary outcome, since there is no determinate connection between the two. For example, which pains would be stripped away in the above example in light of the imaginary outcome?

We must, I believe, view the responsibility record as *two-dimensional*: the basic account which consists of those outcomes for which one is presumptively responsible (roughly: the foreseen impact of autonomous choices) and an imaginary impact account that can be relevant for adjusting the normative significance (e.g., blameworthiness or praiseworthiness) of the basic account. Both are relevant, but they cannot simply be merged (since imaginary impacts are not real). In the case of the doctor above, the additional pain of the patient would be in the basic account, but the imaginary benefits would be in the imaginary account. Both would be taken into consideration in determining, for example, the appropriate reactive attitudes towards the doctor.

This is a complex and important issue that deserves further analysis. Here I shall assume this two-dimensional conception of responsibility and turn to a slightly different question. Instead of focusing on what is in an agent's responsibility record, let us consider the *normative significance* of such records. Obviously, the normative significance of a record depends on the normative perspective: morality, prudence, or something else. To make the topic manageable, I shall focus on prudential assessment. Moreover, I shall focus on assessing the (prudential) *value* of the agent's record rather, for example, than determining what reactive attitudes towards the agent are appropriate with respect to the record. (The two are presumably intimately related, but I leave this open.) I focus on the prudential value of an agent's record because this is what certain standard versions of brute luck egalitarianism (and related sufficiency and priority views)

do. They distinguish between the prudential value (e.g., wellbeing) of outcomes for which an agent is responsible and the prudential value for which the agent is not responsible.<sup>12</sup> They seek to equalize the prudential value of outcomes for which agents are not responsible.

I shall thus focus on the assessment of the prudential value of an agent's record for the purpose of determining what portion of the total outcome value is attributable to the agent's choice (and, for example, thus not subject to any equalization requirement). My question concerns how imaginary outcomes affect this assessment.

Consider, for example a case where an agent intentionally undergoes a painful medical procedure because she falsely imagines that it will cure a disease she has. The foreseen impact of her choice, let us suppose, *reduces* her wellbeing by 10 units (relative to the relevant baseline). Relative to the agent's beliefs, however, her choice will *increase* her wellbeing by 20 units because of the imaginary benefits of the operation. For what impact on wellbeing is the agent responsible? The actual impact was to reduce wellbeing by 10 units and this impact was entirely foreseen by the agent. Is she thus responsible for the 10-unit reduction in wellbeing? Or is she responsible somehow for a 20-unit increase in wellbeing (even though that was not the actual impact)?

I shall suggest that the presence of imaginary outcomes can, relative to the value of the foreseen impact, *shrink* the value for which the agent is responsible but never *magnify* the value. Shrinking involves moving the value closer to zero (e.g., from 4 to 3 or from -4 to -3), whereas magnifying involves moving the value further away from zero (e.g., from 4 to 5 or from -4 to -5). Imaginary outcomes, that is, can mitigate, but never enhance, responsibility for value.

The analysis that follows is very speculative. The topic of how imaginary outcomes affect the value for which the agent is responsible has not, to the best of my knowledge, been

systematically examined before. My aim is therefore very modest. I merely seek to sketch one seemingly promising solution to the problem of imaginary outcomes. My goal is, not to defend this solution but rather, to say enough so that we have a clear picture of the problem and of one possible solution. Even if the proposed solution is quite mistaken (and I won't be surprised if it is), my hope is that thinking about it will help us understand the problem.

#### 4. Imaginary Impact and Responsibility for Prudential Value

How do imaginary outcomes affect responsibility for prudential value? A natural thought is to appeal to the prudential value of the foreseen impact of the choice that the agent *would* have made, had she not had her false beliefs about the imaginary outcome. It would be implausible, however, to hold that agents are responsible for the impact of such hypothetical choices, since this would violate the causal condition on responsibility. It would imply that agents are responsible for impacts that they did not produce (the hypothetical impacts). Instead, the appeal to the hypothetical impact must be a necessary (rather than sufficient) condition. The idea must be that agents are responsible for their actual causal impacts only if they would also have produced them in the absence of their false beliefs. Call this *the hypothetical impact requirement*. I shall now identify some problems with this requirement.

There is, of course, a general worry about whether sufficiently robust counterfactuals are well defined in most cases, but I shall not pursue this issue here. Another worry is whether counterfactuals can be made to track the right kinds of considerations for responsibility. For example, it might be that an agent who, through weakness of will, chose her believed *second best* alternative would have chosen her *first best* alternative, if she had not had the false beliefs about the imaginary impact. If this is possible, then the appeal to the counterfactual seems

inappropriate.

Another problem is that the hypothetical impact requirement seems to reduce responsibility too much. Suppose, for example, that we are asking whether an agent is responsible for getting wet in the rain. Let us suppose that she had a choice between staying home, going out with an umbrella, and going out without an umbrella, and that she chose to go out without an umbrella knowing that she would get wet. Finally, suppose that the only reason that she went out in the rain was because she falsely believed that the house was going to collapse (an imaginary impact). If she had not had this false belief (or any other one concerning imaginary impacts), she would not have gone out and would not have gotten wet. According to the hypothetical impact requirement, the agent is not responsible for getting wet. This, I claim, is implausible. To see this, consider a second agent in exactly the same situation (including the same beliefs) who chooses to go out in the rain *with* an umbrella. Unlike the first agent, he makes the effort to find the umbrella so as to avoid getting wet. Clearly, in at least some situations of this sort, the first agent bears some responsibility for getting wet. Although false beliefs about imaginary outcomes may *reduce* an agent's responsibility for getting wet in such situations, they not completely *eliminate* it, as the hypothetical impact requirement holds.

The issue is complex, but the above considerations at least motivate considering an alternative approach. I shall now sketch an approach on which, instead of appealing to the outcome of some counterfactual choice, one compares the foreseen value impact of a choice with its believed (foreseen and imaginary) value impact.<sup>13</sup> The latter, I shall suggest, can mitigate but never enhance responsibility for the foreseen value impact.

This analysis appeals to the notions of the foreseen value impact and of the believed value impact. In a given choice situation, the agent's beliefs about the values of the outcomes of

her choices (including imaginary outcomes) and other factors determine her disposition to make various choices. Given this disposition, we can calculate two baseline values for the choice situation. One, the *foreseen value baseline*, is the value of the choice situation based on the values of the foreseen outcomes of the choices. The other, the *believed value baseline*, is the value of the choice situation based on the values of the believed outcomes of the choices. They are both based on the same choice disposition (which is based on the agent's actual beliefs) and represent the value of being in the choice situation. They differ solely with respect to the values that they associate with each choice: the value of the foreseen outcomes or that of the believed outcomes (which includes imaginary outcomes). The *foreseen value impact* of a choice is the difference between the value of its foreseen outcome and the foreseen value baseline. The *believed value impact* of a choice is the difference between the value of its believed outcome and the believed value baseline.

Consider, then, the following example. An agent confronts three possible choices, C1, C2, and C3. She (correctly) foresees the outcome of each choice (there is no unforeseen outcome), but falsely anticipates certain imaginary outcomes as well.

Choice	Outcome		
	Foreseen	Imaginary	Believed
C1 (.7)	10	10	20
C2 (.2)	0	14	14
C3 (.1)	9	1	10

For this example, I shall assume that the foreseen value and believed value baselines are

determined on the basis of the expected value of the outcomes, given the agent's choice disposition (given her beliefs). For illustration, we shall assume that the agent has a 70% disposition to choose C1 (which has the highest believed value), a 20% disposition to choose C2, and a 10% disposition to choose C3 (which has the lowest believed value). The basic idea is that the agent is more disposed to choose those options that she believes have greater value, but she is not a perfect chooser: she may make mistakes, suffer from weakness of the will, and so on. The exact probability numbers used here do not matter. Moreover, I believe that my main points remain correct even if the baselines are not determined on the basis of the probabilistic dispositions of agents to choose. I have made the assumption to provide a concrete example of how the baseline might be determined.<sup>14</sup>

Given the assumption that baseline values are determined on the basis of expected values, and given the above probabilities, the foreseen value baseline is 7.9 ( $= .7 \times 10 + .2 \times 0 + .1 \times 9$ ) and the believed value baseline is 17.8 ( $= .7 \times 20 + .2 \times 14 + .1 \times 10$ ). Note that the same probabilities are used for both calculations. This is because we are appealing to the agent's actual choice disposition, given her false beliefs. The only difference is whether we appeal to the (correct) foreseen value or the (mistaken in this case) believed value.

Let us now ask, for each choice, what value the agent is responsible for if she makes that choice. If she makes C1, the foreseen value outcome is 10 and thus the foreseen value impact is 2.1 ( $= 10 - 7.9$ ). The believed value outcome is 20 and thus the believed value impact is 2.2 ( $20 - 17.8$ ). The two impacts have the same sign (both positive) and this reflects the fact that both the foreseen outcome and the believed outcome are improvements over their respective baselines. Relative to the agent's beliefs, the agent is improving things, and this is also true relative to the (actual) foreseen impact. I suggest that in this case, the agent is responsible for the foreseen value

impact—and not the greater believed value impact. Because we are concerned with broad responsibility, we are interested only in the extent to which the actual impact is attributable to the agent's choice. In this case, we attribute the entire foreseen value impact to the agent. It would not make sense to attribute more than this. (Imagine a situation like the above but where the agent falsely imagines even greater benefits. This would not increase her responsibility for the actual impact.)

More generally, where the foreseen value impact and the believed value impact *have the same sign* (e.g., both positive) and the absolute size of the believed value impact is *greater* or equal, the imaginary outcome has, I suggest, no effect on responsibility for value. The agent is simply responsible for the foreseen value impact (which is smaller or equal in absolute size). The imaginary outcome in this case neither enhances nor mitigates responsibility.

Consider now C2. Its foreseen value outcome is 0 and thus the foreseen value impact is -7.9 ( $= 0 - 7.9$ ). The believed value outcome is 14 and thus the believed value impact is -3.8 ( $14 - 17.8$ ). From both perspectives, the outcome of C2 is worse than the relevant baseline. In a case such as this, where the believed value impact and the foreseen value impact have the same sign, but where the believed value impact is *smaller* in absolute size than the foreseen value impact (-3.8 vs. -7.9 in this case), the agent is responsible, I suggest, for the believed value impact (and not the foreseen impact, as I suggested for C1). Because of the false beliefs about imaginary outcomes, the agent was not fully aware of the impact of her choice. The false beliefs do not entirely eliminate responsibility, since even relative to her beliefs her choice (C2) had a negative value impact. They do, however, mitigate responsibility by limiting responsibility to the (smaller in absolute size) believed value impact.

Finally, consider C3. If the agent chooses C3, the foreseen outcome is 9 and thus the

foreseen value impact is 1.1 (= 9-7.9). The believed outcome is 10 and thus the believed value impact is -7.8 (10-17.8). Here, unlike, the previous two cases, the value of the believed value impact does not have the same sign as the foreseen value impact (-7.8 vs. 1.1). The agent is improving things from the perspective of the (actual) foreseen value impact but making things worse from the perspective of the (partially mistaken) believed value impact. Here, the agent is not, I claim, responsible for any value. Her beliefs about the outcome of her choice are too disconnected from reality for her to be responsible for the impact of her choice.

In sum, my suggestion is that agents are (1) not responsible for any value where the foreseen value impact and the believed value impact have different signs (e.g., one positive and one negative), and (2) responsible for the *smaller* absolute value where they have the same. Imaginary impacts, that is, can mitigate but never enhance responsibility for value.

On this proposed account, responsibility for the foreseen value impact is limited—not by appealing to the outcomes of a hypothetical choice but rather—by appealing to the believed value impact of the choice in question. I won't attempt to defend this view, since I am by no means convinced that it is correct even in broad outline. My goal is simply to outline a seemingly promising alternative to hypothetical accounts.

## 5. Conclusion

We have been addressing the topic of broad (agent-) responsibility for outcomes. This is an important topic for any moral view—such as brute luck egalitarianism—that holds that moral accountability for outcomes depends in part on whether the agent is responsible for the outcome. Unlike narrow responsibility, broad responsibility is concerned with responsibility for outcomes outside of the agent's head. For simplicity, we have focused throughout on cases where the agent



is not responsible for her beliefs (e.g., because she is making her first choice). We have been asking how false beliefs affect broad responsibility for outcomes.

False beliefs generate at least three relevant phenomena. One is that they change the agent's disposition to choose, and this, I have assumed, affects the baseline relative to which responsibility for outcomes is assessed. I have had nothing to say here about this important topic. We have simply assumed that the *impacts* of choices are measured relative to such a baseline. A second phenomenon generated by false beliefs is the existence of unforeseen outcomes—actual outcomes that the agent did not foresee. Agents are clearly not responsible for such outcomes and I have had little to say about this topic. A third phenomenon generated by false beliefs is the existence of imaginary outcomes—outcomes anticipated by the agent but with no basis in reality. This is the topic on which I have focused.

With respect to responsibility for outcomes, I have suggested that we need to think of it as two-dimensional: the foreseen impact account and the imaginary account. The two cannot be merged into one account, since broad responsibility is concerned with actual impacts and imaginary impacts are not real. Instead, the foreseen impact account gives us presumptive responsibility and the imaginary account provides the basis for adjusting such responsibility.

The above concerns the issue of what is in an agent's "record". That leaves open how the normative significance of a record with imaginary impact is determined. The normative significance might concern the appropriate reactive attitudes towards the agent with respect to the record or it might concern the value of the record (the value for which the agent is responsible). Such assessments will vary by normative perspective. A record, for example, may be praiseworthy from the perspective of prudence but not from that of morality. Rather than attempting to address the general issues of normative significance of an agent's record, I have

focused on just one: that of determining the prudential value of an agent's record (the value attributable to the agent). This issue is central to standard forms of brute luck egalitarianism (which exclude such value from equalization).

I have suggested that imaginary outcomes (1) eliminate all responsibility for value when the believed (foreseen plus imaginary) value impact has the opposite sign from the foreseen value impact, (2) reduce responsibility for value when the believed value impact has the same sign as the foreseen value impact but has a lower absolute value, and (3) do not affect responsibility for value when the believed value impact has the same sign as the foreseen value impact but has a higher absolute value. My weaker and more general claim is that imaginary impacts sometimes eliminate responsibility for value, sometimes partially reduce it, sometimes do not affect it, and never enhance it.

The proposed analysis focused on responsibility for prudential value. I believe, but shall not argue, that the account is just as plausible as an account of responsibility for value of any sort (e.g., moral value). Moreover, although I have not addressed the question of how imaginary impacts affect the appropriate reactive attitudes (e.g., for morality), a natural extension is to say that they are determined by the (e.g., moral) *value* for which the agent is responsible. I leave this, however, as an open question.

All of this is highly speculative. My goal has been to open up an issue for further investigation rather than to resolve it conclusively.<sup>15</sup>

## Notes

---

<sup>1</sup> For example, T.M. Scanlon, “The Significance of Choice,” in *The Tanner Lectures on Human Value VIII*, ed. Sterling McMurrin (Cambridge: Cambridge University Press, 1988), pp. 151-216; T.M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1998), ch. 6.

<sup>2</sup> See, for example, Joel Feinberg, “Action and Responsibility,” *Philosophy in America* edited by Max Black (London: George Allen & Unwin, 1965), pp. 134-60 [Reprinted in Joel Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton: Princeton University Press, 1970), pp. 119-51]; Jonathan Glover, *Responsibility* (London: Routledge and Kegan Paul, 1970), p. 44; Thomas Nagel, “Moral Luck,” *Proceedings of the Aristotelian Society*, Supplementary Volume 50 (1976): 137-51; Michael J. Zimmerman, *An Essay on Moral Responsibility* (Rowman and Littlefield, 1988); and Ishtiyaque Haji, *Moral Appraisability: Puzzles, Proposals, and Perplexities* (Oxford University Press, 1998).

<sup>3</sup> On the connection between agent-responsibility and reactive attitudes, see, for example, Peter Strawson, “Freedom and Resentment,” *Proceedings of the British Academy* 48 (1962): 187-211; and R. Jay Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press, 1994).

<sup>4</sup> Agent-responsibility for outcomes is similar to what is called “moral responsibility” for outcomes, except that it builds in no requirement that the agent be responsive to moral reasons. See, for example, John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1999) and Michael J. Zimmerman, *An Essay on Moral Responsibility* (Totowa: Rowman and Littlefield, 1988).

---

<sup>5</sup> See for example, Peter Vallentyne, “Brute Luck Equality and Desert”, in *Desert and Justice*, edited by Serena Olsaretti (Oxford University Press, 2003): 169-185. This view is also held by Michael J. Zimmerman, “Taking Luck Seriously,” *Journal of Philosophy*, 99 (2002): 553-576; Michael J. Zimmerman, “Luck and Responsibility,” *Ethics* 97 (1987): 374-86; and Judith Jarvis Thomson, “Morality and Bad Luck,” *Metaphilosophy* 20 (1989): 203-221 [Reprinted in *Moral Luck*, edited by Daniel Statman (Albany: State University of New York, 1993)].

<sup>6</sup> Some might argue that agents are responsible only for their intended outcomes and not for merely believed outcomes. For simplicity, I focus on the full believed outcome, but I allow that this may need to be further narrowed down.

<sup>7</sup> Broad responsibility is at least roughly what is called “indirect responsibility” in Michael Zimmerman, *An Essay on Moral Responsibility*.

<sup>8</sup> It is also the notion of responsibility that Fischer and Ravizza address in *Responsibility and Control*. They call it an “externalist” account (pp. 252-253).

<sup>9</sup> For more on how one can be responsible for false beliefs, see Michael J. Zimmerman, “Moral Responsibility and Ignorance,” *Ethics*, Vol. 107 (1997): 410-426; Carl, Ginet, “The Epistemic Requirements for Moral Responsibility,” *Philosophical Perspectives* 14 (2000): 267-77; Gideon Rosen, “Culpability and Ignorance,” *Proceedings of the Aristotelian Society* 103 (2003): 61-84; and Gideon Rosen, “Skepticism about Moral Responsibility,” *Philosophical Perspectives*, 18 (2004): 295-313.

<sup>10</sup> I develop an account of the baseline in Peter Vallentyne, “Brute Luck and Responsibility”, *Politics, Philosophy & Economics* 7 (2008): 57-80.

<sup>11</sup> See “Brute Luck and Responsibility”.

---

<sup>12</sup> Here and throughout I follow Susan Hurley use “brute luck” in the thin sense of “not responsible for”. See ch. 4 of Susan Hurley, *Justice, Luck, and Knowledge* (Cambridge, MA: Harvard University Press, 2003).

<sup>13</sup> I compare the value of the believed (foreseen and imaginary) impact with that of the foreseen impact rather than just comparing the value of imaginary impact with zero, because value may be holistic (so that the value of the believed impact is not the sum of the values of the foreseen and imaginary impacts).

<sup>14</sup> I defend the view that the value of the baseline is the expected value of the agent’s choices in “Brute Luck and Responsibility”.

<sup>15</sup> For helpful comments, I thank Dick Arneson, Carl Knight, Kasper Lippert-Rasmussen, Martin O’Neill, Shlomi Segall, Zofia Stemplowska, Hillel Steiner, Andrew Williams, and Jo Wolff.