

# “Free Will” is Vague

Santiago Amaya  
Universidad de los Andes

In this paper, I argue that free will is vague. Although the claim can be made more specific, the simplest formulation is that the predicate “is free,” as applied to human decisions and actions, is vague. Thus, even if there are clear cases of free and unfree actions, there will be plenty of borderline cases. With respect to them, the answer to the question “Is this action or decision free?” is simply indeterminate.

I begin with a clarification of the thesis, followed by a two-stage argument. I argue, first, that free will is a matter of degrees and, second, that there are no sharp boundaries separating free and non-free actions and decisions. Then, I apply the thesis to manipulation arguments, which many have thought to support skepticism regarding compatibilist free will, showing why considerations of vagueness detract from their apparent strength. In the end, I conclude by sketching some further consequences of the thesis.

## 1. The Thesis

Discussions of vagueness and free will have occurred quite independently from each other. In fact, only recently questions regarding the semantics of the terms in the free will debate have begun to surface (Vargas, 2015; Deery, 2019). So, it might be useful to begin with a clarification of the thesis.

### 1.1. Vagueness

Philosophers have long recognized that many natural language expressions are vague (Williamson, 1994). “Bald,” “tall,” and other humdrum terms are commonly used as illustrations of the phenomenon. Yet, as some have noted, some vague expressions are at the heart of well-known philosophical discussions, for example, “knowledge” (Sorensen, 1987) and “vagueness” (Hu, 2017).

Among philosophers of language, there are disagreements with respect to how vagueness ought to be understood. Fortunately, we do not need to settle on a specific account here. It is widely agreed across the different accounts that vague expressions share two interrelated features: they admit of borderline cases and are susceptible to sorites.

*Borderline cases* are those in which it is indeterminate whether a given term applies to an individual. Some men are clearly bald, others are not, yet some are borderline: they seem to be neither bald nor not bald. This is why many classical logicians have abhorred vagueness. Statements attributing a vague predicate to a borderline case are *prima facie* threats to the principle of bi-valence.

Some cases of seeming indetermination are *not* due to vagueness: if they can ultimately be solved by conceptual analysis or empirical inquiry. Yet, neither method will serve to determine whether someone who, say, is borderline bald, is in fact bald. In Roy Sorensen's (2018) terms, borderline cases are inquiry resistant. There is no further inquiry that could be carried out to settle the question of whether the predicate applies to them, unless the inquiry is meant to redefine the extension of the term.

Now, in so far as vague terms admit of borderline cases, statements predicating them can generate *sorites paradoxes*. This is a consequence of these terms not having a precise extension. In general, a sorites is the argument form exemplified by the conjunction of the following claims:

[1] 1 is a small number

[2] For any natural number, if the number is small, then adding 1 to that number will result in a small number

[3] Hence, all natural numbers are small

Put more abstractly, sorites arguments lead from a base premise to a conclusion, via a generalization principle. Characteristically, the base premise seems intuitively true, the conclusion patently false, and the generalization principle hard to deny. Suppose you deny the truth of [2]. Which is the small number such that adding 1 to it will result in a large number?

Notably, the paradoxical nature of the sorites can be brought out, even if one is agnostic regarding the conclusion of the argument. This shows that their paradoxical nature depends little on one's intuitions about particular cases. In short, every *ascending* sorites can be mimicked to give rise to a corresponding *descending* sorites, one whose conclusion is the negation of the former and whose generalization premise is construed by taking an antonym of the vague term (Hyde & Raffman, 2018). To illustrate:

[1\*] 1 trillion is large number

[2\*] For any natural, if the number is large, then subtracting 1 to that number will result in a large number

[3\*] Hence, all natural numbers are large

## 1.2. Free will

As used in philosophical discussions, “free will” is a *term of art*. Although non-philosophers have beliefs and make claims about free will, the relation between those beliefs and assertions and the philosophical term “is free” is rather complicated and does not presently concern us. The claim to be defended here is primarily about the term as used in philosophical debates.

Even with this focus, free will is a rather indeterminate notion. But, the lack of precise meaning is not necessarily a reflection of its vagueness. Let us, for the sake of clarity, distinguish two sources of imprecision that are definitely *not* due to vagueness. First, free will is invoked in different evaluative domains, which suggests that the conditions for the application of the predicate might vary as a function of context. Theorists, for instance, often analyze free will as a necessary condition for *moral responsibility*, in particular, moral accountability. Yet, the term also captures what many think are conditions for other kinds of evaluations: autonomy, personal value, etc. (see Kane, 1996: 81-88).

Equally important, contemporary discussion of free will has occurred against the backdrop of the Consequence Argument and Frankfurt cases. Out of this discussion, philosophers have distinguished two conceptions of freedom, providing yet another

reason to doubt of the univocality of the term (Timpe, 2013). There is, on the one hand, the *leeway* conception of free will, which the Consequence Argument primarily targets: free will as the ability to do otherwise. There is, on the other hand, the *source* conception that Frankfurt cases seek to rehabilitate: free will as a condition of authorship.

Behind these distinctions there are issues worthy of consideration. Does free will mean the same in each of these evaluative domains? How does leeway and source free will relate each other? In what follows, I remain neutral on those matters. For ease of exposition, I will use “free will” to refer to a condition on moral responsibility, although the argument I present can be generalized to other notions. Also, I shall alternate from leeway to source understandings of the term to exemplify the more general points at different junctures of the argument.

### **1.3. The claim**

Although free will can be invoked in different evaluative contexts, one can easily make more precise one’s talk by means of simple specifications, for instance, as I did in the immediately preceding paragraph. Yet, even after the context is specified and the term is disambiguated, there will be underdetermined cases. That is the idea behind the claim that “is free” is vague. Unless one precisifies the term in ways that depart from established usage or regiment it, there will be borderline cases.

Hints to this idea can be found in the literature, at least, if one reads in-between the lines. But because philosophers have been mostly concerned by the “bugbears” of determinism, they have not explicitly discussed or thematized the point at any length.<sup>1</sup> Peter van Inwagen provides a nice example of this attitude. As he argues, free will, which he considers to require the ability to do otherwise, is exercised in deliberate choices. But deliberate choices, according to him, are relatively rare in human life. For the most part, our decisions and our actions are a consequence of automatisms and

---

<sup>1</sup> One exception worth mentioning is Ichinose (2008), who quickly notes that free will is vague and discusses how the vagueness of free will serves to push the idea that vagueness is an ontic phenomenon. McKenna (2012: 10) also notes the possibility of vagueness of the predicate “is free” as applied to persons.

internalized routines, which he views as lying at the opposite side of a continuum that goes from the free to the non-free:

“The light turns green, and the driver, his higher faculties wholly given over to thoughts of revenge or lunch or the Chinese Remainder Theorem puts his car into gear and proceeds with his journey. Did he do something called ‘making a choice between proceeding and not proceeding’? Presumably not: the whole thing was too automatic. The young public official, unexpectedly and for the first time, is offered a bribe, more money than he has ever thought of having, in return for an unambiguous betrayal of the public trust. After sweating for thirty second, he takes the money. Did he make a choice? Of course. Between these two extremes lie all sort of cases, and it is probably not possible to draw a sharp line between making a choice and acting automatically.” (van Inwagen, 1989: 404)

The last sentence of the passage merits attention. If there are borderline cases of choices and automatic forms of making up one’s mind, are there also borderline cases of free will?

Unfortunately, van Inwagen does not press the issue any further; he goes on to try to identify the kind of situations that clearly implicate deliberate choices. So, we cannot be sure what his answer to this question would be. Regardless, setting aside his specific account of the nature of choices, which might be idiosyncratic within the free will debate, his attitude towards the vagueness of “is free” is symptomatic of much of the literature. A focus on the question about the existence of free will has prevented philosophers from asking how free we are and how sharp the difference is between free and non-free decisions and actions. They have mostly distinguished different senses of “free will” and proceeded to give an account of seemingly paradigmatic cases of free or non-free actions along their preferred understanding of the term.

## **2. Argument**

Having clarified the terms and the underlying thought behind the thesis, I now turn to the argument. Let us consider its two main premises.

## 2.1. Premise 1: free will is a matter of degrees

As mentioned earlier, judging by their use of the predicate “is free,” it would seem as though philosophers treat free will as an all or nothing affair. But, for the reasons sketched then, this can be charitably interpreted as “shallow” talk, prompted by the salience of skepticism in the debate.<sup>2</sup>

It is possible, however, to look at what conditions are often claimed, not just to be necessary, but to actually *ground* free will (Sartorio 2016). Interestingly, many of the candidates here involve scalar notions, which provides a good reason to think that free will comes in degrees. If a decision or action is free in virtue of certain conditions and those conditions can be met to different degrees, then it would seem that one’s decisions or actions can be free to different degrees.<sup>3</sup>

Think about the notion of *leeway* freedom. According to philosophers who understand free will along these lines, “is free” is applicable to agents who have the ability to decide or act otherwise. Plausibly, like any other ability, this is one that each of us could have to varying degrees. Just to mention one possibility, doing otherwise might be hard for those who have compulsive urges, easy for those who have well-developed skills for self-control, and something in between for most of us.

Now, as skeptics have argued, it is possible that no human action or decision is ever free. But that possibility, as we discuss below, is generally meant to be realized due to some contingent non-linguistic feature of the world.<sup>4</sup> There could be worlds in which

---

<sup>2</sup> But see Coates & Swenson (2013), Nelkin (2016), and Kaiserman (2020) who recently discuss scalar notions of free will.

<sup>3</sup> The argument could also be made in terms of defeaters if partial absences are not included among grounds. Roughly, if a decision or action *ceases* to be free in virtue of conditions that can obtain to different degrees, then it would seem that one’s decisions or actions can be gradually less free. See section 3 below.

<sup>4</sup> According to a form of skepticism, the conditions for the application of the predicate “is free” are simply self-contradictory (Strawson 1994; Smilansky 2001). That accusation would certainly be problematic for the proposal advanced here, because it would mean the predicate “is free” lacks an extension altogether. Here, I assume that we can meaningfully discuss whether the conditions for freedom obtain in our world. So, I set aside this radical form of skepticism.

human agents are free in the sense in which we actually use the term. In those worlds, the ability for doing otherwise can be differentially possessed by individuals who are better or less equipped to resist the pressures to behave in some specific way. Some of them, accordingly, could act or decide more freely than others.

We can make the point about degrees more concrete by inspecting a standard way of developing the notion of leeway freedom favored by many *libertarians*. According to them, to the extent that causation by prior events (together with the laws of nature, etc.) necessitates its effects, causal determination precludes alternative possibilities. Yet, they contend, causal determination does not pervade our world, specifically, they believe that we have good reasons to think that some of our choices are not causally necessitated by other events. This is what opens for agents the possibility to do otherwise

Actually, libertarians typically do not even require that our choices lack any causes at all. What matters for a choice to be free is not that it has no preceding causes, but that those causes do not fully settle the choice for the agent. It matters, on their view, that there is some significant room for the agent to make a different choice. As Mark Balaguer puts it:

“At one end of the spectrum, which option is chosen is wholly uncaused...; at the other end of the spectrum, which option is chosen is causally determined, i.e, nomologically necessitated by the state of the universe together with causal laws; and in between, there is a continuum of possible cases... I think that libertarians should respond to this by allowing that (conceptually speaking) there can be different degrees of L-freedom.”  
(Balaguer, 2004: 384)

Unlike libertarians, *compatibilists* need not postulate degrees of indetermination. Many of them, at least those persuaded by Frankfurt cases, oppose the leeway conception of free will and instead believe that what is key for decisions and actions to be free is to be brought about in ways that can be attributed to their agents. The point here is stronger. Their conception of free will is not just one that can easily be made compatible with a scalar approach; it *encourages* such an approach (Nelkin, 2016; Brink 2021, ch 3)

Consider, reasons-responsiveness accounts. According to this version of compatibilism, acting and deciding freely is a matter of being able to recognize moral reasons and regulate one's behavior accordingly. The presumption is not that only morally acceptable actions and decisions count as free. Rather, the idea is that actions are free because they result from a *capacity to respond to a spectrum of reasons*, a spectrum wide enough for the actions to be a function of the moral sense of the agent and, hence, attributable to her.

It is precisely here that the issue of degrees comes in. To make room for individual free actions and decisions that are somewhat unreasonable, reason-responsive compatibilists need to ground freedom in the agent's capacity to respond to reasons. But, in addition, to exclude undesirable forms of necessitation, they also need to exempt from the domain of freedom agents who, due to some *psychological incapacity*, might only occasionally respond to certain reasons. To make this last move, a common avenue is to postulate that the relevant capacity comes in degrees. Simply, some agents are not reason-responsive enough for their actions and decisions to be regarded as free.

Witness, in this regard, the lessons that Michael McKenna derives from the comparison between two hypothetical agents: Handy, the compulsive hand washer, who can occasionally respond to certain reasons, and Dandy, the handwasher who cares about looking clean but at times fails to wash his dirty hands:

“The cases of Handy and Dandy suggest two related features of a credible reasons- responsive theory. First, to understand freedom in terms of reasons-responsiveness involves an agent's control in relation to a spectrum of potential but non-actual reasons, and this spectrum comes in *degrees*. Second, the *actual causal processes* issuing in free action must involve the agent's reasons-responsive resources.” (McKenna 2016: 26)

Obviously, the appeal to reason-responsiveness is only one way for compatibilists to develop their position. But it is easy to see how the point generalizes to other varieties of compatibilism. In general, compatibilists need to offer grounds for distinguishing, among causally determined actions and decisions, those that count as free and those that don't. And it is reasonable to think that the difference between the former and the latter



would not be categorical, say, like the differences between actions and mere behaviors, but a difference in degrees.

Just to drive the point home, think about the problems faced by the early compatibilist proposal of Frankfurt (1971). Having categorically distinguished the free and non-free by the presence of certain second-order desires, Frankfurt shortly after amended his theory due to the possibility that agents could be somewhat alienated from some of their second-desires. Following him, many deep-self theorist nowadays prefer to invoke, not a hierarchy of desires, but a set of attitudinal complexes, which ground characteristics, such as identification (Frankfurt 1992) or self-expression (Sripada 2016), etc., all of which were evidently scalar.

With this, let's set aside for the moment the discussion of specific accounts and return to the general argument. Assuming that free will is a matter of degrees, we shall now consider whether there is a sharp boundary separating free and non-free actions and decisions. I shall argue that there is no sharp boundary, which will allow us to see why free will is vague. This will also allow us to discuss an objection to the present argument.

## **2.2. Premise 2: there is no sharp boundary separating the free and the non-free**

Any vague term can be precisified in ways that allow for a sharp separation of the items within and outside of its extension. But precisifications of vague terms necessarily involve *some degree of stipulation*. The predicate "is a child," for instance, does not have a sharp boundary. Yet, legal systems can define a precise age at which someone can be treated as an adult only because they are allowed to stipulate on pragmatic grounds.

Precisifications of the predicate "is free" are possible and might be useful. But any attempt to precisify it will involve some significant stipulation. In brief, given the standard conditions that free actions and decisions are typically required to meet, there is no principled way to mark the precise degree to which these conditions need to be met. Some conditions come with a precise metric, but *no* point in that metric seems useful to capture the difference between the free and the non-free. Conditions that

promise to capture the distinction, by contrast, are *not* plausibly fitted with a precise metric.

As an example of the former, think back to the idea that freedom, understood as the ability to do otherwise, features some degree of indetermination. Assume, for the reasons discussed above, that full causal indetermination (together with certain epistemic and psychological requirements) is sufficient but not necessary for free will. What degree of indetermination separates free and non-free actions and decisions?

*Objective probabilities* are one way to try to answer this question. Indeed, it is a natural avenue, given how libertarianism is often discussed in the free will debate. Objective probabilities seem to account for the libertarian phenomenology of decision-making, in particular, of one's reasons having varying degrees of strength but not forcing one to act one way or the other. Also, they seem to allow libertarian free will to conform to a certain "scientifically respectable" understanding of non-deterministic causation (O'Connor, 2009, Vicens, 2016).

Clearly, from this perspective, any free action and decision has to have prior to its occurrence a probability less than 1. Yet, for reasons familiar from well-known objections to libertarianism, it is not evidently clear which lesser objective probability is right to mark the differences here. If the probability of a putative free decision or action were to approach 0.5 then, as some have argued (Van Inwagen 2000: 14-15), a natural response would be to conclude that the action or decision is not so much the result of the agent's freedom but of *randomness* or chance. On the other hand, if the probability approaches any of the extremes (0 or 1), then it seems that this version of libertarianism actually confers freedom to agents even when they do not have a *feasible* opportunity of doing otherwise (see Pereboom 2014: 12; Nelkin 2016: 359).

For present purposes, we do not need to agree with the objections from which these claims are taken. The problem is not whether the threshold for freedom is set here or there, but that there really is no principled way of deciding where the boundary ought to be drawn. In short, sharp variations in probabilistic degrees do not seem to correspond to the significant differences that are supposed to be captured by labelling some actions and decisions as free and others as non-free. In fact, to the extent that such labeling has

implications for how agents are appraised and treated, small percentual differences in indetermination seem off the mark.

John Fischer eloquently makes this claim in connection to free will understood as a condition of moral responsibility. He says:

“I do not think that our status as agents should hang on a thread—that is should depend on whether natural laws have associated with them (say) probabilities of 0.99 or 1.0. In my view that sort of empirical difference should not *make a difference* to our moral responsibility.” (Fischer 2012: 3-4)

Now, Fischer intends his remark as a dismissal of incompatibilism about free will, at least as a motivation for endorsing his semicompatibilism. Except for possibly the skeptic, who would want our status as free agents to hang on a thread? But the point does not really count against libertarianism. Perhaps free will requires indetermination and, as suggested above, indetermination ought to be cashed in terms of numeric objective probabilities. Yet, provided that one does not saddle libertarianism with a task that cannot be fulfilled, namely, finding a sharp difference between the free and non-free, there isn't much of an objection here. In fact, similar considerations can be raised with respect to other commonly accepted positions in the debate.

Consider, once again, reason-responsiveness accounts of free will. Contrary to differences in probabilistic degrees, there might be good considerations for thinking that there are sharp differences in reason-responsiveness among possible targets of the predicate “is free”. And, one might think that those differences create a meaningful distinction between those within and outside of the extension of the predicate. If, for instance, prudential reasons and moral reasons are of a different kind, lack of responsiveness to reasons of the latter type might mark a sharp difference between those agents whose actions and decisions can be free and those without this power.

The problem, however, is that these distinctions do not come with a metric that can provide the right boundary. They point to sets of considerations that are sharply distinct, say, the prudential vs. the moral. But the distinctions are orthogonal to the issues raised

above in connection to the scalar notion of reason-responsiveness. Thus, whereas a person who responds to both the prudential and the moral is arguably more responsive than the person who only responds to one of them, the complicated cases for compatibilists, as we discussed earlier, arise because reason-responsiveness comes in degrees even *within* these domains. Sticking to the moral domain, it is not sufficient for a free action or decision to be made by an agent who can respond to some moral reasons. It matters that the agent can respond to *moral reasons* in a sufficiently robust way.

The problem becomes clearer if we turn instead to proposals that allow for degrees *within* domains of reasons. Some reason-responsiveness theorists, for instance, invoke possible worlds to explicate the modal aspects involved in this notion (see Vargas, 2013, ch. 7, Brink 2021, ch 3). As they see it, degrees of reason-responsiveness can cash out in terms of the proportion of relevantly different possible worlds in which the agent acts in appropriate ways. The more possible worlds in which an agent responds to moral reasons, the more reason-responsive the agent is.

By considering sets and subsets of possible worlds, these proposals can make sense of graded reason-responsive capacities. The drawback, as can be easily anticipated now, is that there is no form to settle in a precise way how large the set of possible worlds in which the agent decides or acts accordingly ought to be in order for her to be free, except by sheer stipulation. In fact, the very idea of trying to identify the precise cardinality of the subsets of possible worlds that matter for freedom seems hard to motivate.

Again, we should not view this as an indication of a deep problem for reasons-responsiveness accounts (but see Herdova & Kearns 2017, p. 174). Rather it is an indication that these theories should not be saddled with the impossible task of trying to find a sharp boundary between the free and the non-free. Once we recognize that free will comes in degrees, we shall also recognize that comparison in degrees of reasons-responsiveness will be irremediably vague. That is, we should accept that there is no sharp boundary between the free and the non-free.

### **2.3 Graded notion?**

It could be objected that the present argument rests on a problematic inference.<sup>5</sup> It is true, as stated above, that the grounds for free will can obtain to a greater or lesser degree. But, according to the objection, it does not follow that free will comes in degrees. Even if an action or decision is free provided that some conditions are met to certain degree, all actions and decisions that are free are equally so (Fischer, 2006: 233).

No doubt there are plenty of everyday predicates with this kind of profile, for instance, “being accepted to college X.” Although the conditions for acceptance might be met to varying degrees, at the final stage of the process every applicant is either accepted or not, and nobody is “more accepted” than anyone else. Even though the conditions for applying the predicate come in degree, the predicate itself is not scalar.

Obviously, for the proposal to work here, it should be possible to find the right threshold over which the freedom conditions can be considered as met. But, as we have seen, at least if you consider the standard ways in which “is free” is explicated in the current debate, no sharp appropriate boundary can be found. Which means that even if there is a threshold that separates the free and the non-free, it would have to be a vague threshold. Regardless, the predicate “is free” turns out to be vague.

As an illustration, consider the predicate “is an adult.” Presumably, everyone who counts as an adult, counts equally so. But if we think that a child ceases to be a child when they become an adult, the vagueness of “is a child” spills over “is an adult.” Even though there is a threshold, to the extent that the latter is vague, there will cases to which neither predicate definitely applies. Thus, the predicate is vague even if it is not graded.

We can now return to the overall argument and formulate it in more precisely:

---

<sup>5</sup> It might also be asked whether considerations about vagueness are extendable to all possible conceptions of free will. Perhaps, they are not, but aiming for such generality is unnecessary for present purposes. The goal here is to discuss whether the term “is free” as used in the current debate is vague, so inquiring about merely possible uses is a far more ambitious goal. If there are possible approaches to free will that explicitly exclude a graded notion, I venture to say that they are somewhat revisionist of the term.

The conditions that ground freedom can be met to lesser or greater degrees. But there is no precise threshold that marks the difference between free and non-free actions and decision. This means that either there is no exact minimum point at which actions and decisions become less or more, but definitely, free or that there is a continuum of cases to which the predicate “is free” does not determinately apply or fail to apply. Either way, the predicate “is free” is vague.

### **3. Skepticism**

Issues about the degrees and vagueness of “is free” have significant consequences for the free will debate. In this section I consider how the vagueness of free will illuminates some arguments that skeptics have widely deployed to cast doubt on compatibilist freedom, the so-called manipulation arguments.

#### **3.1. Manipulation arguments**

Skepticism about free will comes in a variety of forms. But perhaps the most common form of skepticism involves the claim that “is free” is *contingently* a predicate lacking extension. According to it, there is no member in the extension of the predicate in the actual world because of the pervasive presence of certain defeaters of freedom in it. But there are other possible worlds in which the cluster of conditions that prevent decisions and actions from being free in this world are absent.

Standard manipulation arguments seek to motivate skepticism of this kind with respect to compatibilist free will, in particular, desert-based free will. They are to meant to show that in a world in which determinism holds the predicate “is free” lacks extension. The reason is that plausible candidates for agency and/or desert in deterministic worlds are not significantly different from manipulated agents, who by hypothesis are incapable of making free decisions or acting freely.

Manipulation arguments, however, implicitly rely on the vagueness of free will. In rough outline, they take us from clear cases of freedom being defeated by manipulation to border-line cases. And, from there, they move to cases of agents who meet

compatibilist conditions for freedom, but live in deterministic worlds. The intended conclusion is that these agents are not free, given that there are *no sharp differences* between these and the cases consider earlier.

Let me make this point more precisely. Manipulation arguments are often described as having the following overall form (Mele 2008; McKenna 2008):

[4]  $S_1$  is not free

[5] Concerning freedom relevant features, there is no significant difference between  $S_1$  and  $S_n$ , where  $S_n$  is an agent who meets all compatibilist conditions for free will.

[6]  $S_n$  is not free

Clearly, this is a simplification. Although the inference from [4] to [6] is characteristic of manipulation arguments, most of the action in them takes place in establishing [5]. In fact, the way [5] is usually defended obscures how vagueness actually plays a role in them. Often these arguments are discussed as though the conclusion were reached by directly comparing a determined and a manipulated agent, when in reality that “direct” comparison is secured by considering variations in the manipulated case meant to address putative compatibilist responses (see Mele 1995 and Sartorio 2016, ch.5). Thus, what seems to be a direct argument is actually one that involves a set of multiple comparisons.<sup>6</sup>

It is here that the vagueness of “is free” works its way; and it’s also where these arguments ultimately fail. To the extent that we compare similar cases of agents with insignificant differences regarding their freedom (or absence of it), we seem to be allowed to generalize from one case to the next in whatever judgements we make with respect to their freedom. But the appearance is deceptive, once we recognize that there is no sharp threshold separating the free and the non-free. Because the generalization from manipulated to non-manipulated agents is infected by vagueness, the arguments do not actually prove the intended conclusion.

---

<sup>6</sup> Similarly, Sartorio (2016: 157-158) distinguishes, in this regard, simple and bridge versions of manipulation arguments, and notes that simple versions of it tend to be embedded in complex bridge versions.

### 3.2. 4-case vagueness

To make my case concrete, we can focus on Derk Pereboom's (2001, 2014) 4-case argument, which has the further advantage of having a relatively transparent structure. In it we are asked to consider four different scenarios in which Plum, an agent who meets all compatibilist conditions for free will, winds up deciding to kill someone else. In the first two scenarios a team of scientists manipulates Plum's brain to produce the murderous decision. In the third, he is led to the decision by a set training practices of his community. Finally, in the fourth scenario his decision results from a deterministic history of the kind that compatibilists claim does not defeat freedom.

According to Pereboom, in the first two scenarios Plum does not decide freely. The third scenario, he claims, is not significantly different from the second—the community practices play in it the defeating role of the intervention of the scientific team in the first two scenarios. So, the same intuition ought to hold there. The final scenario is not to Pereboom's eyes significantly different from the third, which leads him to his skepticism about compatibilist free will. In short, in all four cases Plum does not decide freely because his decision is causally determined by factors beyond his control.

Various philosophers have tried resisting the argument by seeking some significant difference between the scenarios that could, in turn, lead to differences in attributions of freedom (Demetriou 2010; Deery & Nahmias (2017)). But, as McKenna (2008, 2014: 468) has rightly observed, the strategy is “dialectically unstable:” the cases can easily be reformulated (or intermediate ones can be added) to secure a smooth application of the *generalization* strategy up to the last scenario where no manipulation is present. This is why he recommends compatibilists to offer Pereboom a hard-line reply. Instead of trying to find significant differences among the cases, they should embrace their similarity. And rather than beginning from Pereboom's first scenario, they should start from the last and move down to the first to argue that, even then, Plum makes a free decision.

Although McKenna's response is insightful, given our discussion so far, we should have some hesitation. In fact, there is reason to suspect of the overall dialectic behind the 4 case-argument and the hard-line response. Suppose, as the dialectic presupposes, that



free decisions are those that can be traced to the agent in ways that makes her deserve certain forms of treatment. Now assume, in the light of the argument of the previous section, that freedom conditions come in degree and the threshold for freedom with respect them is vague. Then, to the extent that Pereboom and McKenna are right about the cases being similar enough, what seems to allow the generalization from each scenario to the next is ultimately the vagueness of free will. Although each case is not exactly alike as the next, the differences in the way in which the agent is externally influenced seems not significant for the purposes of applying the predicate “is free.”

Consider the upward transitions that Pereboom favors. He generalizes from the first to the second scenario because at-the-moment and temporally-distant manipulation do not seem different enough from each other. At the same time, McKenna seems equally entitled to move downward from the fourth to the third case. If determination by the laws of nature does not defeat agential source, the influence of one’s community in one’s upbringing does not seem to subtract enough of it to place the decision beyond the agent’s free will. The same, of course, happens with respect to the transition from the second to the third scenario, and vice versa.

In order to avoid the stalemate, Pereboom and McKenna embark on a discussion about the appropriate epistemic attitude to approach the argument and one’s intuitions with respect to each of the cases. McKenna insists that his reply only requires being agnostic about the truth of compatibilism; Pereboom insists that we should be ready to revise whichever attitudes we initially had as we move down the cases. Yet, if the preceding remarks are on the right track, this further discussion is a distraction. It would be like asking which attitude should one have towards the size of the natural numbers to decide which, of the *ascending* [1-3] and *descending sorites* [1\*-3\*] described above, should one embrace.

Now, Pereboom (2001: 116) quickly dismisses the possibility of treating the manipulation argument as a sorites, on the grounds that as one moves through the cases “there is not divergence at all in factors that could possibly make a difference for moral responsibility.” But the generality of his claim begs the question against the present argument. It is true that between the first and the last case there are no differences in the degree of determination. Still, as one moves along the cases there are graded differences

in how the agent is externally influenced and those differences matter for assessing whether, with respect to freedom, certain kinds of external interventions and determinism are on a par.

Let me explain. As presented above, the argument for the vagueness of “is free” focuses on the *grounds* that make true attributions of the predicate. But we can focus instead on the conditions that *defeat* applications of it and run the exactly the same kind of reasoning. The conditions that defeat freedom can obtain to more or less degree; with respect to them there is no precise threshold marking the difference between the free and the non-free. Thus, even if some conditions that are central to ground free will are determinately met (e.g. the agent being definitely reason-responsive), application of the predicate in certain cases might be indeterminate because of the degree to which freedom defeaters obtain. Because of that, the generalization from similar cases on a continuum, as reflection on the sorites shows, might not preserve truth.

To illustrate, Andrew Latham and Hannah Tierney (2021) have recently shown that attributions of free will are sensitive to the scope of external interventions. As a matter of psychological generalization, an agent who is *singly* determined tends to be considered less free than the agent who is equally determined but lives in a world where *every agent is equally determined*. This, according to them, explains why the first two cases in Pereboom’s argument look like no-freedom cases. But it also explains the grip of the compatibilist intuition that fuels McKenna’s response. Presumably, which external interventions count as defeaters of freedom is determined against some standard of normalcy shaped by how most people’s decisions and actions are believed to come about. That’s why it matters whether the external interventions happen singly or universally.

What happens, however, when we move between these extreme cases? What if the agent is part of a community of determined agents, as the agent in Pereboom’s third case might well be? Is there a precise number at which that community, if it were to grow, would become free? These are clearly rhetorical questions. Likely, the threshold here is a vague: there is no precise number at which population proportions become the norm. So, there is no exact point at which external determination ceases to be seen as defeating or being compatible with free will.

This is obviously only one illustration of how issues of vagueness can creep in here. Likely, it is one against which a skeptic who distrusts the correctness of normal practices and assumptions would push.<sup>7</sup> But others could be cited here.<sup>8</sup> The 4-case argument is supposed to work against any compatibilist understanding of free will, so we cannot peg the vagueness on which it revolves to a concrete account of freedom or what defeats it. But this does not make the problem go away. Even if we cannot commit to a specific explanation of how the argument works, the hypothesis does point to its overall limitations. Given that the threshold that distinguishes the free and the non-free is vague, generalizing from apparently similar cases is a bad strategy, for it is not truth-preserving.

To be clear, this conclusion is premised on the reading that Pereboom and McKenna offer of the dialectic of the argument. Thus, philosophers who think that the 4-case argument fails because significant differences can be *identified* among the cases will admittedly not find the point made here compelling. It won't work against those hardliners who would like insist that all cases are *strictly identical* with respect to presence of defeaters of freedom. But, if Pereboom and McKenna are right that the cases can be arranged and organized to allow generalization from one to the next, and these generalizations turn out to range over defeaters that can be more or less present, then we have a good reason to suspect that the argument is irremediably flawed.

### 3.3. Soft-line vagueness

---

<sup>7</sup> Pereboom (2014: 76), for instance, contends that single and momentary external interventions are compatible with freedom: behaving grumpily because one's team loses an important game. So, this isn't a factor that should make a difference here. But this is not a strong response to Latham and Tierney's point, as the neuro/psychological mechanisms by which news of the defeat affect one's mood negatively are quite normal and widespread, unlike the neuroscientific intervention of case 1.

<sup>8</sup> Another illustration is provided by Usher's (2018) proposal, where a gradation of more and less robust causes (those whose effects depend more or less on background conditions) is used to account for the logic of manipulation arguments. In particular, as one moves from case 1 to case 4 in Pereboom's argument, external interventions become less robust, causally speaking, which gradually permits the agent to become more the locus of control.

So far, I've focused on McKenna's hard-line response and brushed away soft-line responses, which insist on their being significant differences among the cases. But, perhaps, this dismissal was a bit too quick. To finish the discussion of manipulation arguments, I want to discuss an option open to soft-liners.

As mentioned at the outset, vagueness has been regarded as a threat to the principle of bi-valence. In response, epistemicists have argued that vagueness ultimately reduces to a form of ignorance. According to this line, vague predicates have a definite extension, but their exact extension is something that competent speakers of the language do not and cannot know. Thus, as far as epistemicists are concerned, sorites arguments are valid. The problem with them is that the generalization principle ([2] and [2\*] above) is false. There is, so they argue, a small natural number such that adding 1 to it makes it big.

Epistemicism has definitely an air of *mysterianism*. Why would speakers be irremediably ignorant of the extension of a predicate in their own language? Some epistemicists think that no answer is needed: knowledge needs an explanation but not the absence of knowledge (Sorensen, 1988). But perhaps the most widespread answer, championed by Timothy Williamson (1994), is that ignorance of the exact extension of vague terms is a consequence of certain margin-of-error principles. In short, knowing precisely what is in the extension of a vague term would be incompatible with the kind of safety constitutive of knowledge.

Soft-line responses can take a page of this book to undercut manipulation arguments. In true epistemicist spirit, they can resist the generalization from manipulated agents to causally determined agents, on the grounds that there is, as one moves away from intentional-at-the-moment-manipulation, a point in which causal determination ceases to be a defeater of freedom. Further, following Williamson's line, they can argue that we cannot know what that exact point is because it would cross margins of error in ways that are incompatible with knowledge attributions.

Here's one way of developing this kind of response. Typically, manipulation arguments begin with cases where the mechanisms of causal determination, despite being part of some outlandish scenario, are clearly defined. Yet, as they move away from these cases,

the causal stories of how determination happens invariably get more abstract and less detailed. What is, for instance, the causal route by which community practices shape one to have a certain motivational structure? This is unfortunate, especially if, as compatibilists would argue, variations in the causal history of an action or decision can make a difference with respect to its freedom.

Regardless, under these circumstances, intuitions with respect to which of a series of small variations mark the boundaries of freedom would seem to be unreliable. And, if this is true, it provides the soft-line respondent with a non-mysterious explanation as to why we cannot know where the sharp separation of the free and the non-free lies. Simply put, whatever intuitions we have here cannot become knowledge because they lie at the margins of error. They are intuitions about cases in which false belief is a live possibility.

This is, obviously, only a possible response for soft-line detractors of manipulation arguments to adopt. They can adopt other alternative responses, depending on what account of vagueness they prefer to endorse. What is important for the present argument is not so much the precise response to manipulation argument but the consequences that considerations of vagueness have for their apparent strength.

#### **4. Concluding remarks**

I have claimed that free will is vague and have provided an argument for that thesis. Further, given the vagueness of free will, I have argued that manipulation arguments should not be considered as decisive to evaluate the truth of compatibilism. This is just one interesting consequence of the thesis. But there are others worth exploring.

For example, to the extent that free will is a condition for moral responsibility, the vagueness of “is free” likely spills over “is morally responsible.” Thus, even though there are cases of wrongdoing where the agent is determinately guilty and some where she isn’t, there will be also be cases that are simply indeterminate: the agent does something wrong but does not have a proper excuse for it. Without some dose of stipulation, we won’t be able to say whether the agent is morally responsible or not for them.

Stipulation, of course, need not be a bad thing. There might be good reasons to revise our use of the terms “is free” and “is responsible” to include in their extension, or to exclude from it, cases that would otherwise be considered borderline. Perhaps, by doing so, we can change some of our practices and habits, so as to avoid preventable but not necessarily culpable forms of wrongdoing. Or we can become more charitable to those who have done wrong but who might not be clearly at fault.

What we should avoid, however, as the discussion of manipulation arguments illustrates, is the use of borderline cases as springboards to generalize to other cases that are not borderline. Whereas the resulting arguments might not quite look like sorites, by moving through seemingly small differences, we might end up reasoning in ways that are not truth preserving.<sup>9</sup>

### References

- Balaguer, M. (2004). A coherent, naturalistic, and plausible formulation of libertarian free will. *Noûs*, 38(3), 379–406.
- Brink, D. (2021). *Fair opportunity and responsibility*. Oxford: Clarendon Press.
- Coates, J., & Swenson, P. (2013). Reasons-responsiveness and degrees of responsibility. *Philosophical Studies*, 165(2), 629–645.
- Deery, O. (2021). Free actions as a natural kind. *Synthese*, 198(1), 823–843.
- Deery, O., & Nahmias, E. (2017). Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood. *Philosophical Studies*, 174(5), 1255–1276.
- Demetriou, K. (2010). The soft-line solution to Pereboom’s four-case argument. *Australasian Journal of Philosophy*, 88(4), 595–617.
- Fischer, J. (2006). *My way: Essays on moral responsibility*. New York: Oxford University Press.
- Fischer, J. (2011). *Deep control: Essays on free will and value*. New York: Oxford University Press

---

<sup>9</sup> I would like to thank Carolina Sartorio, Derk Pereboom, Manuel Vargas, and Michael McKenna for helpful comments to earlier drafts, as well as participants of the III Conference on Philosophy and Neuroscience at the Universidad Federal do Rio de Janeiro.,

- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5-20.
- Herdova, Ma., & Kearns, S. (2017). This is a Tricky Situation: Situationism and Reasons-Responsiveness. *The Journal of Ethics*, 21(2), 151–183.
- Hu, I. (2017). ‘vague’ at higher orders. *Mind*, 126, 1189–1216.
- Hyde, D., & Raffman, D. (2018). Sorites Paradox. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. URL=  
<https://plato.stanford.edu/archives/sum2018/entries/sorites-paradox/>.
- Ichinose, M. (2008) Vagueness of Free Will. *Proceedings of the XXII World Congress of Philosophy*, 15, 53-58.
- Kaiserman, A. (2020). Reasons-sensitivity and degrees of free will. *Philosophy and Phenomenological Research*. doi:10.1111/phpr.12738
- Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.
- Latham, A., & Tierney, H. (2021). The four-case argument and the existential/universal effect. *Erkenntnis*. doi:10.1007/s10670-021-00458-x
- McKenna, M. (2008). A hard-line reply to Pereboom’s four-case manipulation argument. *Philosophy and Phenomenological Research*, 77(1), 142–159.
- McKenna, M. (2012). *Conversation and Responsibility*. Oxford: Oxford University Press.
- McKenna, M. (2014). Resisting the manipulation argument: A hard-liner takes it on the chin. *Philosophy and Phenomenological Research*, 89(2), 467–484.
- McKenna, M. (2016). Reason-responsive theories of freedom, In K. Timpe, M. Griffiths and N. Kevy (Eds), *The Routledge Companion to Free Will* (pp. 27–40), New York: Routledge
- Mele, A. (1995). *Autonomous agents: From self control to autonomy*. Oxford: Oxford University Press.
- Mele, A. (2008). Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics*, 12(3–4), 263–286.
- Nelkin, D. (2016). Difficulty and degrees of moral praiseworthiness and blameworthiness. *Noûs*, 50(2), 356–378.
- O’Connor, T. (2009). Degrees of freedom. *Philosophical Explorations*, 12(2), 119–125.
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford,: Oxford University Press.

- Sartorio, C. (2019). *Causation and free will*. Oxford: Oxford University Press.
- Smilansky, S. (2001). Free will: From nature to illusion. *Proceedings of the Aristotelian Society*, 101(1), 71–95.
- Sorensen, R. (2018). Vagueness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). URL=  
<https://plato.stanford.edu/archives/sum2018/entries/vagueness/>
- Sorensen, R. (1987). The vagueness of knowledge. *Canadian Journal of Philosophy*, 17(4), 767–804.
- Sorensen, R. (1988). *Blindspots*. Oxford: Oxford University Press.
- Sripada, C. (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies*, 173(5), 1203–1232.
- Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, 75(1–2), 5–24.
- Timpe, K. (2013). *Free will in philosophical theology*. New York, NY: Continuum Publishing Corporation.
- Usher, M. (2020) Agency, Teleological Control, and Robust Causation. *Philosophy and Phenomenological Research*, 100(2), 302-324.
- van Inwagen, P. (1989). When is the will free? *Philosophical Perspectives*, 3, 399-422.
- van Inwagen, P. (2000). Free will remains a mystery. *Philosophical Perspectives*, 14, 1–20.
- Vargas, M. (2015). *Building Better Beings: A theory of moral responsibility*. Oxford: Oxford University Press.
- Vicens, L. (2016). Free will and theological determinism. In K. Timpe, M. Griffiths and N. Kevy (Eds), *The Routledge Companion to Free Will* (pp. 535–544), New York: Routledge
- Williamson, T. (1994). *Vagueness*. Oxford, England: Routledge.