

Cet article a été téléchargé sur le site de la revue Ithaque :
www.revueithaque.org



Ithaque : Revue de philosophie de l'Université de Montréal

Pour plus de détails sur les dates de parution et comment soumettre un article, veuillez consulter le site de la revue : <http://www.revueithaque.org>

Pour citer cet article : **Banville, F.-I. (2012) « Implementing Revisionism: Assessing a Revisionist Theory of Moral Responsibility », *Ithaque*, 10, p. 115-135.**

URL : <http://www.revueithaque.org/fichiers/Ithaque10/Banville.pdf>

Cet article est publié sous licence Creative Commons « Paternité + Pas d'utilisation commerciale + Partage à l'identique » :
<https://creativecommons.org/licenses/by-nc-sa/2.5/ca/deed.fr>



Implementing revisionism : Assessing a revisionist theory of moral responsibility

Frédéric-Ismaël Banville*

Abstract

The aim of this paper is to examine a particular substantive theory among others in the set of “revisionist” theories of moral responsibility, namely, Manuel Vargas’ version of the moral influence account of the justification of responsibility-specific practices. Moderate revisionism, which Vargas (2005) endorses, advocates a clear distinction between descriptive and normative questions, which enables a naturalistically plausible account of responsibility that does not jeopardize the normative aspect. However, while Vargas provides a useful framework for thinking about revisionism, I argue that despite its initial appeal, an actual revisionist theory does not seem to track as closely as we would like what I call the “meta-theory” of revisionism, viz. what Vargas defines as the features of moderate revisionism. Outlining these differences enables the formulation of observations about 1) the role of revisionist approaches for theorizing about moral responsibility and 2) how revisionism can be integrated with scientifically informed approaches.

Thinking about responsibility (almost always) leads to the well-trodden paths of the problem of free will and to the question of how it can be reconciled with determinism. Intuitively, knowing whether free will is compatible with determinism is of the utmost importance : if determinism is true and free will does not exist, then we cannot hold anyone (morally) responsible. Accordingly, theories of moral responsibility have been categorized on the basis of their take on that particular interrogation. Either determinism is true and we do not have free will (hard determinism) or, determinism is not true, or at

* L'auteur est étudiant à la maîtrise en philosophie (Université du Québec à Montréal).

least does not hold for us as humans (various strands of libertarianism) or, determinism is compatible with free will (compatibilism).

Revisionist approaches differ from these traditional theories of responsibility and as Manuel Vargas¹ argues, at least one particular species, moderate revisionism, is of particular interest. First, it is immune to (or at the very least, it is less vulnerable to) the usual criticisms levelled at these approaches (viz. the “familiar argument”, discussed in section 1.2), and, second, it has the virtue of bringing to the forefront concerns shared by most if not all theories of moral responsibility, viz. telling us how we should treat each other under the relevant circumstances. On this view, a theory of responsibility is primarily a *practical* theory. I will not go into much details here (section 1 will take care of that), but it is important to note that this focus on the practical aspect comes at a price. Traditionally, normative theories of responsibility are justified because they are anchored to a metaphysical property of responsibility : our concept of responsibility is appropriate because it refers to the right kind of thing. On an account of the type Vargas puts forward, not only is the traditional metaphysical question of the compatibility of determinism with responsibility (or free will) sidestepped, but the metaphysical nature of responsibility becomes partially irrelevant. Provided revisionism can effectively and justifiably limit the scope of the discussion to, for instance, the justification of various norms and practices, the question is about *how* we justify our responsibility practices, until such time as we decide metaphysical questions are to be addressed.

The aim of this paper is to examine a particular substantive theory among others Vargas develops. Vargas advocates moderate revisionism, a type of approach which, when understood correctly, provides an alternative to the traditional theories about responsibility that purportedly preserves what is *really* important for theories of responsibility, while doing away with the pitfalls of the metaphysical debate about free will and determinism. However, while Vargas provides a useful framework for thinking about revisionism, I argue that despite its initial appeal, an actual revisionist theory does not

¹ Vargas, M. (2005), “The Revisionist’s Guide to Responsibility”.

seem to track as closely as we would like what I call the “meta-theory” of revisionism, viz. what Vargas defines as the features of moderate revisionism. To defend this point, a detailed outline of the particular type of revisionism Vargas endorses is necessary and will be provided in section 1. Section 2 will deal with Vargas’ positive proposal. Finally, I will show that there is a discrepancy between the level of independence the various components enjoy towards one another in the meta-theory and in the implemented version.

1. What revisionism is, and what sort of a revisionist Vargas is

In this first section I provide a comprehensive outline of what I call the “meta-theory” of revisionism as developed by Vargas. Making clear what moderate revisionism entails at the meta-theoretical level will make it possible to highlight the discrepancies between what revisionism *should* be, and *what it is* once we implement it. A caveat is in order though : I will be focusing on a specific type of revisionism, one that Vargas calls “moderate revisionism” because it is the sort of revisionism he explicitly endorses².

1.1. *What is revisionism ?*

Broadly speaking, revisionism is the idea that some aspect of our responsibility-characteristic practices, attitudes, and concepts are in need of revision³. To make this definition a bit clearer, let us take a look at what theories of responsibility are usually supposed to be doing, and what they are about. First, it is relatively uncontroversial to suppose that theories of responsibility typically have to answer three questions : the metaphysical question (about the nature of

² This is because, in part, moderate revisionism is mostly immune to the “familiar argument” (see Vargas, M. (2005), “The Revisionist’s Guide to Responsibility” for a detailed discussion of the “familiar argument”), which posits that any form of revision of our concept of responsibility necessarily entails the elimination of our responsibility-characteristic practices. This sort of argument is often directed at strong revisionism, because it tends towards eliminativism about our practices, concepts and attitudes.

³ Vargas, M. (2005), “The Revisionist’s Guide to Responsibility”, p. 399.

responsibility, viz. *what is* responsibility), the descriptive question (about what our responsibility-characteristic practices, attitudes and concepts *are*), and the normative question (about what those things *ought to be*)⁴. Moreover, a theory of responsibility typically has to work with three things : 1) psychological dispositions (roughly what Strawson⁵ calls *reactive attitudes*), 2) responsibility-characteristic practices (rewarding, for instance) and 3) our folk-concept of responsibility (cluster of beliefs and concepts we hold about responsibility)⁶.

It is clear that the two last questions call for different projects, one *descriptive*, and the other *normative*⁷. What makes an approach revisionist is the fact it prescribes something else than what it diagnoses. This might seem, *prima facie*, to be strange, but think simply of Dennett's particular brand of "compatibilism"⁸, especially in his 1973 paper⁹. We might very well be completely determined (viz. mechanism¹⁰ is true), but we are still justified in viewing each other as morally responsible agents, because it is the only way we have to

⁴ Vargas, M. (2005), "The Revisionist's Guide to Responsibility", p. 402.

⁵ Strawson, P. F. (1960), "Freedom and Resentment", *Proceedings of the British Academy*.

⁶ Vargas, M. (2005), "The Revisionist's Guide to Responsibility", p. 402.

⁷ Vargas (2005) uses *diagnostic* and *prescriptive*, probably to minimize the risk of confusing the normative aspect of a theory of responsibility with a theory of normative ethics. I use the terms interchangeably, unless otherwise noted.

⁸ It is unclear whether Dennett is a compatibilist or a revisionist, and it surely is not an easy question to answer. Vargas (2005) proposes that the Dennett of *Freedom Evolves* (2003) can be read both ways and argues in favour of a revisionist reading, while McKenna (2009) considers Dennett a "multiple viewpoint compatibilist", in reference to Dennett's reliance upon the intentional stance. Considering Dennett's account exhibits clear revisionist features, I will take for granted a revisionist reading, since even if it turns out he is a compatibilist, it seems safe to assert he would be a revisionist sort of compatibilist.

⁹ Dennett, D. C. (1973), "Mechanism and Responsibility".

¹⁰ Here, mechanism refers to the idea that by virtue of our being part of a natural order that is determined by a set of laws and so on, human beings are just as much a predictable mechanism as is a plant, and we have no free will, no agency of the sort we intuitively think we would have (Dennett, D. C. (1973), "Mechanism and Responsibility", p. 253).

interact with each other in the relevant contexts. This serves as a good illustration of what Vargas means when saying a revisionist theory is one where there is a divergence between normative and descriptive projects. This particular sort of revisionism is what Vargas calls “paradigmatic revisionism” (a revisionism which puts forward a separation between normative and prescriptive, as opposed to other forms of revisionism), which can be broken down into three types.

The first of those is weak revisionism which is revisionism of our beliefs about some elements of a theory of responsibility. In short, weak revisionism asserts we need to revise our understanding of responsibility, but does not entail any sort of revision of the concept itself. Strong revisionism, on the other hand, entails the elimination of some or all of our concepts, attitudes, or practices, generally as a consequence of scepticism about responsibility itself (whatever it may be)¹¹. Between those two positions lies moderate revisionism which is the type of revisionism Vargas endorses.

1.2. Moderate revisionism

Moderate revisionism is the idea that our folk concept of responsibility is inadequate until it has undergone some measure of revision. This is what Vargas calls moderate conceptual revisionism. If a theory identifies this sort of conceptual error, but maintains that the property of responsibility exists and does not need to be revised, then the theory is of the “error-success” type : there is a conceptual error but there is no property error. On the other hand, a theory could assert that the error extends to the property of responsibility, in which case it is of the “error-error” type. I will not discuss error-success revisionism further, as it comes with a set of particularly complex problems, especially if one intends to preserve some form of indeterminist or libertarian account of responsibility¹². Error-error revisionism, however, does come with its own set of constraints.

¹¹ Vargas, M. (2005), “The Revisionist's Guide to Responsibility”, p. 408-409.

¹² The main problem for an error-success account is that it has to explain how it is that there *is* a property of responsibility but that we do not have the right concept to refer to it. One could devise a causal theory of reference,

Before moving on to the constraints of error-error revisionism, there is something to be said about how moderate revisionism evades the “familiar argument”, which in essence posits that any revision of our concept of responsibility or of the property of responsibility entails the elimination of our practices, attitudes and so on. Recall the distinction drawn in 1.1 between the different things a theory of responsibility is about : psychological dispositions, practices, and folk concept. Revisionism can be about any of those things while leaving two of them untouched, or eliminate one of those things, revise another and leave the third alone. Revisionism, thus, can vary between categories. Revisionism can also vary within categories. Conceptual revisionism, for instance, could simply be about the freedom condition of responsibility, while leaving other parts of the conceptual apparatus untouched, or it could revise only a small subset of our practices. This means, then, that moderate revisionism, even of the error-error type, is never committed to wholesale elimination, thus dodging the “familiar argument” bullet. However, error-error revisionism must comply with two very important constraints.

The first constraint, which is applicable to any moderately revisionist theory, is the plausibility constraint. Revisionist theories gain some ground by being open to the possibility of a normative account that is different from the descriptive account, which enables a naturalistically plausible descriptive account of responsibility that does *not* jeopardize the normative part of the theory in the way determinism¹³ is traditionally thought to. However, the normative account has to be plausible : a theory of responsibility that prescribes something that is sociologically impossible or simply implausible in light of our generic scientific knowledge about ourselves and our

perhaps, but the fact remains that, *prima facie*, attempting to preserve the property of responsibility, or stating that it is not in need of revision, imposes considerable demands on a theory, which might be more complicated to meet than if we simply state that the property of responsibility is not what we thought it was and needs to be revised (see Vargas, M. (2005), “The Revisionist's Guide to Responsibility”, p. 415-417).

¹³ “Naturalistically plausible” is taken to entail, to some extent, the truth of a certain sort of determinism, in the sense that any naturalistically plausible descriptive account will take humans as part of nature and subject to the same causal laws as anything else in the set of things that exist in nature.

psychology is simply inadequate. In a sense, the descriptive account is constrained by naturalistic plausibility, and the normative account is constrained by the descriptive.

The second constraint is the “warrant” constraint and is specific to error-error revisionism. Given that this type of revisionism holds that the property of responsibility itself needs to be revised (or eliminated), the error-error theorist is committed to two things. First, there is no metaphysical property of responsibility or such a property needs to be revised¹⁴. Second, if our practices and attitudes are to be preserved *or* if we want to get to a (normatively) justified prescriptive project, we need something to act as an anchoring point for whatever practices we deem worthy of being prescribed¹⁵. In other words, the warrant constraint states that if there is no property of responsibility, or if such a property is not what we thought it was, we need another property, X, such that if we use X as an anchoring point for our responsibility-characteristic practices, they will be well-justified¹⁶.

The question of how we are to justify our practices is precisely the one which Vargas tries to answer with the proposal we analyze in the following section. By way of concluding remark for this section, it seems appropriate to emphasize that what has been discussed so far is the meta-theory of revisionism, a sort of general framework that outlines the structure of a certain type of revisionist theory.

2. Beyond the framework : the justification of our practices

A revisionist account of the sort Vargas advocates, because it seeks to ground our understanding of responsibility practices, attitudes and so on in a naturalistically plausible way, cannot rely on things like agential causation¹⁷. Such a theory has to provide a set of responsibility norms, the justification for these norms, as well as an

¹⁴ Vargas, M. (2005), “The Revisionist's Guide to Responsibility”, p. 413-441.

¹⁵ *Ibid.*, p. 413-441.

¹⁶ *Ibid.*, p. 413-441.

¹⁷ Chisholm, R. (2007), “La liberté humaine et le moi”, p. 325-341.

account of the candidates to which those norms apply¹⁸. Our concern here is with a *normative* account of responsibility, which requires a plausible and warranted account of the justification of our practices, and so I will focus on this part of Vargas' proposal, which relies on moral influence theory (hereafter MI), though I will mention the important aspects of the descriptive theory.

2.1. Moral influence

MI theory is part of the traditional consequentialist model of responsibility and asserts that our responsibility-characteristic attitudes and practices are justified because they influence, in a forward-looking way, agents to act in a socially desirable fashion¹⁹. This particular version, however, is beset by considerable problems (discussed in 2.2). Vargas' aim is to salvage the insight at the core of MI : responsibility-characteristic practices derive *part* of their justification from the effect they have on the type of creatures we are²⁰.

To understand Vargas' proposal, it is useful to start by breaking down a “complete” theory of moral responsibility in three distinct sub-theories. First, a theory of moral responsibility must provide an account of responsible agency that makes clear the type of agent to which responsibility norms apply²¹. Second, such a theory needs an account of the justified responsibility norms themselves, which leads to the third component, an account of the justification of these norms²². This last component in turn calls for an account of the *aim* of the responsibility system, of what it is directed at²³. On Vargas' view, all that MI can provide is an account of the justification of

¹⁸ While this might seem to be a mostly descriptive concern, an account of responsible agency needs to include at least some normative components, especially when one must determine what sort of agent can be exempted from responsibility judgements and for what reasons.

¹⁹ Vargas, M. (2008), “Moral Influence, Moral Responsibility”, p. 92.

²⁰ *Ibid.*, p. 91.

²¹ *Ibid.*, p. 96. Vargas also provides such an account in “Building a Better Beast” (manuscript).

²² *Ibid.*, p. 96-97.

²³ *Ibid.*, p. 97.

responsibility norms, and this is the core of his argument. Since the main problem of traditional MI is that it is seen both by proponents and critics as attempting to provide a complete theory of responsibility²⁴, scaling down the scope of MI is the first step in making it workable as a justificatory strategy for our practices, though it is not sufficient.

As well as reducing the scope of MI, one needs to refine the notion. Traditional MI is committed to two things : 1) praise and blame are forward-looking attempts to foster socially desirable action and 2) the justification for praise and blame is partly derived from their effects (this is what Vargas thinks we should preserve from MI)²⁵. The problem here is that the first commitment leads to construing the central thesis of MI as follows : the effects of praise and blame are actually the effects of particular tokenings of praise and blame. Vargas suggests that a better way to conceive of the effects of those practices is to see them as arising at the level of general practices²⁶ : the justification for praise and blame arises from the group-level effects of justified norms that come to be internalized by members of the community.

This requires an account of the aim of the responsibility system, as mentioned above. On Vargas' view, the reason why MI works as a justification for our practices is because the responsibility system's aim is to foster a specific kind of agency, one that is sensitive to moral considerations²⁷. This means that we need to provide an account of responsible agency. Interestingly, both the theory of the justification of responsibility norms *and* the account of the aim of the responsibility system make a certain type of account of moral agency more attractive, one which states that it is dependent upon the presence and normal performance of basic psychological mechanisms (beliefs, intentions, etc.) as well as responsiveness to moral considerations²⁸. I will discuss this feature of Vargas' account in detail

²⁴ Vargas, M. (2008), "Moral Influence, Moral Responsibility", p. 100.

²⁵ *Ibid.*, p. 98.

²⁶ *Ibid.*, p. 98.

²⁷ *Ibid.*, p. 98.

²⁸ *Ibid.*, p. 99.

in section 3. For now, however, we need to look at how those various elements come together.

2.2. *Making MI plausible*

The three main components of Vargas' account of the justification of responsibility norms are as follows : the aim of the responsibility system is to foster a distinctive kind of agency, which is defined by various psychological parameters as well as by sensitivity to moral considerations and our practices are justified because, at the level of general practices, they favour the internalization of responsibility norms. If those three things are true, it is required (and this is a descriptive question) that agents be 1) influenceable in virtue of 2) being responsive to moral considerations²⁹. The result is that our practices are justified when those practices, over time, aid responsible agents to act in ways governed by moral considerations³⁰. I will explain how those components interact.

Two things need to be clarified. First, a leftover from traditional MI is the idea that practices will be effective insofar as they exploit our psychology³¹. The contention is that praise encourages us to repeat a particular course of action, or to view a particular type of behaviour as positive, whereas blame does the opposite. Second, the fact the norms of responsibility are regularly enforced by the aggregation of numerous instances of praise and blame lead agents to *internalize* those norms has an important implication, viz. that if this process of internalization involves *justified* norms, it will result in an agent that tracks and responds to moral considerations and, moreover, that does not need to go through *actual* deliberation each time she poses an action³².

This defuses the common concern towards consequentialist approaches that we either are unable to carry out the calculation (deliberation, in this case) underlying our decisions, or that we simply “don’t work that way”. This could be likened to the idea that most of

²⁹ Vargas, M. (2008), “Moral Influence, Moral Responsibility”, p. 99.

³⁰ *Ibid.*, p. 99.

³¹ *Ibid.*, p. 100.

³² *Ibid.*, p. 100.

Implementing revisionism :
Assessing a revisionist theory of moral responsibility

our decisions are made on the basis of heuristics or rules of thumb³³. An MI account could assert that because of the influence the general practices of praise and blame, we come to internalize the responsibility norms upon which those practices are based, and that this results in our decision-making mechanisms relying on reliable heuristics for deciding what to do in a given type of situation. This is consistent with the role the internalization of norms plays in Vargas' account³⁴. Note that this leaves open the possibility of akrasia : those norms can be overridden, an agent could internalize the wrong norms, or fail to properly internalize the justified norms³⁵.

One concern, however, is that this internalization process depends on complex psychological parameters which are to be taken into account in the descriptive aspect of the theory. As seen in 1.2, the normative project of a revisionist theory of responsibility is constrained by the descriptive account that accompanies it, which in turn is constrained by naturalistic plausibility. I will not discuss this further here but it is important to note that knowledge of our moral psychology becomes not only relevant but crucial if we commit to the justificatory strategy Vargas puts forward. Indeed, MI as a justification of our practices relies on our having the *right kind of responsible agency*. This is the beginning of a concern I will detail in section 3, though it also serves as an illustration of the importance of moral psychology for Vargas' theory.

³³ See for example: Kahneman, D., P. Slovic, and A. Tversky. (1982), *Judgements Under Uncertainty- Heuristics and Biases*.

³⁴ The notion of reliable evolved modules could be part of the descriptive story here. For instance, we could have some form of functionally specialized module that relies on social clues to help us make decisions. This is speculative, but as an example, we could think of Kurzban, Tooby and Cosmides' (2007) discussion of coalitional computation: the clues we use to determine if a given individual is part of our group or of another one are extremely flexible and are, presumably, socially determined. This could provide a basis for an account of the internalization process that gives rise to the sorts of heuristics mentioned earlier by explaining how socially acquired parameters make their way to psychological mechanisms which operate in a largely unconscious and automatic way.

³⁵ Vargas, M. (2008), "Moral Influence, Moral Responsibility", p. 100.

One might wonder how we move from justifying responsibility norms as a whole to justifying tokens of praise or blame. The answer is simply that the general norms provide guidelines to think about various types of situations, and that in any given circumstances, an agent will be able to pragmatically assess whether her blame or praise is justified (or an observer will pragmatically decide if a witnessed tokening of praise or blame is justified). This is because the general norms inform us about cases of a certain type³⁶. One might, of course, demand more than a “mere” pragmatic warrant, but this is not the kind of epistemic certainty we can hope for. As a parallel, think of the other-minds problem. How can I be certain that anyone else has a mind? The (generally accepted) answer is that I cannot *know with absolute certainty* if anyone else has a mind; all I *can* rely on are behavioural correlates of having a mind or of having a conscience³⁷. The same is broadly true of moral judgements, and indeed if we think of everyday moral judgements, we do not rely on extremely demanding epistemic credentials of the type that would be required to be absolutely certain a particular instance of praise or blame is justified³⁸. Moreover, given that we consider those norms justified precisely because they contribute to our having the relevant kind of agency (one that is sensitive to moral considerations), we seem to have good reason to believe that, at least in typical cases covered by the generic norms, our judgements will be justified³⁹. I will discuss atypical cases in section 3.

Two last points before we can sum up the MI theory of justification. First, it must be noted that this account is *modular*: when integrated with different theories of normative ethics, there *will* be a change in the type or content of the moral considerations to which we are supposed to be receptive, but the structure of the account of the *justification* of the norms will remain intact⁴⁰. MI, on Vargas’ view, is an account of how the norms that prescribe right action are

³⁶ Vargas, M. (2008), “Moral Influence, Moral Responsibility”, p. 105.

³⁷ See, among others: Harnad, S. (2003), “Can a Machine be Conscious? How?”, and Hyslop, A. (2009), “Other Minds”.

³⁸ Vargas, M. (2008), “Moral Influence, Moral Responsibility”, p. 105.

³⁹ *Ibid.*, p. 106.

⁴⁰ *Ibid.*, p. 108.

justified. Second, and this is related to the first point, such an account of the justification of responsibility norms is *not* normatively exclusive, viz. moral influence is *not* the only way our practices can be justified. In fact, there is a possibility for additional and potentially overlapping ways in which our practices as well as reactive attitudes can be justified⁴¹.

2.3. Summing up

Before moving on to the final section, it seems appropriate to sum up what has been said so far. The three components of Vargas' account of the justification of responsibility norms are as follows : 1) the aim of the responsibility system is to foster a distinctive kind of agency, 2) this moral agency is defined by various psychological parameters as well as by sensitivity to moral considerations, 3) our practices are justified because at the level of general practices they favour the internalization of responsibility norms. Section 2.2 highlighted four characteristics of Vargas' account that are required for his MI-inspired account of the justification of our practices to be workable. First, because Vargas endorses the second commitment of traditional MI, viz. that practices will be effective insofar as they exploit our psychology, the descriptive account of agency which accompanies the normative account of the justification of the norms must take into consideration the features of our moral psychology, in the sense that for the normative account to be plausible, we need to possess the right kind of agency. This kind of agency has to be receptive to moral considerations as well as enable the process of internalization to take place, since Vargas' theory relies on the internalization of the responsibility-characteristic norms.

Second, Vargas offers a provision for dealing with specific cases. Because we internalize the responsibility norms which underlie our practices, deciding whether a particular instance of praise or blame is justified is akin to using a sort of reliable heuristic when taking other sorts of decision. This leads Vargas to asserting that the best we can hope for in the way of epistemic warrant for our judgements is of a pragmatic nature.

⁴¹ Vargas, M. (2008), "Moral Influence, Moral Responsibility", p. 108.

Third, Vargas makes it clear his account is *modular*, in the sense that the justificatory strategy he puts forward does *not* rely on a specific theory of normative ethics. If it turns out that consequentialism is true, then the norms we internalize will be of a consequentialist nature. If it turns out that deontologism is true, then those norms will be of that type. Regardless of *what* the moral considerations or norms are, the story about how they are justified remains the same. This ties in with the fourth feature, which is that the MI theory of justification is not normatively exclusive : there could be additional and overlapping normative accounts that also provide justification for our practices, and Vargas is clear on that point when he says that “the *justification* for praising and blaming practices derive, *at least in part*, from their effects on creatures like us.⁴²” This means that MI is potentially not the only way in which our practices can be justified.

At this point, we have a clear picture of how a revisionist theory of responsibility is structured and what it is committed to.

3. Concerns, criticisms and perspectives

This final section raises two concerns related to the discrepancy between the prescribed structure of a revisionist theory of responsibility, as outlined in the meta-theoretical account discussed in section 1, and the actual structure of the elements of an instantiated case of open revisionism about responsibility.

3.1. Independence of the normative and descriptive

The first concern I wish to address is internal to Vargas’ account and moderate revisionism in general. Recall that what makes revisionism attractive as a general approach to the problem of moral responsibility is that it allows that the different parts of a theory of responsibility be construed as relatively independent from one another. That is, by clearly separating the metaphysical question from the normative and descriptive questions (which are distinct), it becomes easier to identify what should *really* matter to any theory of

⁴² Vargas, M. (2008), “Moral Influence, Moral Responsibility”, p. 91. Emphasis added.

responsibility, viz. how we should treat each other in the relevant contexts. In his 2005 paper, Vargas makes a strong case for revisionism and indeed convincingly shows that a moderately revisionist approach *can* help us move beyond the traditional disagreements between compatibilists, determinists and libertarians. This is, however, at a meta-theoretical level. In what follows I will try to show that taking a closer look at an openly revisionist theory of the justification of our responsibility-characteristic practices highlights the fact that the independence of normative and descriptive accounts is not as clear as we might think.

As shown in 1.2, the normative project of a revisionist theory of responsibility is constrained by the descriptive account that accompanies it, which is in turn constrained by naturalistic plausibility. This means that any normative account, including an account of the justification of our practices of praise and blame, is not receivable if it is in any way implausible given a descriptive account of, say, responsible agency. In the case of the justification of our responsibility practices, the normative account has two components : the justification of our practices *and* the account of the aim of the responsibility system, which is to foster a certain kind of agency in individuals. However, both the theory of the justification of responsibility norms *and* the account of the aim of the responsibility system make a certain type of moral agency more attractive, one in which responsible agency relies upon the presence and normal performance of a host of basic psychological mechanisms (beliefs, intentions, etc.) as well as responsiveness to moral considerations.

Thus, the descriptive account, which informs us about the kind of agency we *can* have, becomes, in an important way, dependent upon the normative account. Of course, we could conceive of the relationship between normative and descriptive as follows : even though the descriptive account is *suggested* by the normative, the normative is still significantly constrained by the descriptive, because it *depends* on the descriptive being true.

This answer is unsatisfactory, because the normative aspect imposes considerable demands on our moral psychology. For the normative account to hold, we not only need to have the right kind of agency, but we also need to have the psychological capabilities and

functioning required to make this kind of agency plausible. It is worth noting that such a reliance on moral psychology has an upside : if we manage to identify the psychological features required for the right kind of responsible agency, determining when a given agent is to be exempted from the demands of responsibility could become easier. This particular point is speculative, and I suspect even with a good understanding of moral psychology, such questions will not be so easily answered⁴³. The point here is that the components of the revisionist account Vargas puts forward are not as independent as the meta-theoretical account suggests, and that it seems that the dependency and constraining relationship is not only from the descriptive to the normative but also from the normative to the descriptive. Whether this is a crippling problem is another question that I will not discuss here, but I believe what has been said so far warrants taking a closer look at how *actual* revisionist theories respect the parameters outlined by Vargas⁴⁴.

3.2. Typical and atypical cases

The other question I wish to address is that of the pragmatic justification for particular instances of praise or blame. Recall that, on Vargas' view, the best we can hope for is a pragmatically grounded justification that is derived from the general rules of thumb our responsibility system provides⁴⁵. This works well for typical cases, but atypical cases may pose a problem. The problem is not so much that Vargas does not provide a solution to atypical cases (he does). Since we internalize a number of norms that guide our action, every singular case is judged on the basis of those norms, and if we are in the presence of an atypical case (whatever that may be), then we simply are *not* within the purview of the normative account Vargas provides, which is an account of the justification of our responsibility-specific practices *in general*. The problem (and it is not

⁴³ For a detailed discussion of the difficulties posed by mental illness for theorizing on moral responsibility, see: Broome, M. R., L. Bortolotti, and M. Mameli. (2010), "Moral Responsibility and Mental Illness: A Case Study".

⁴⁴ Vargas, M. (2005), "The Revisionist's Guide to Responsibility", p. 399-429.

⁴⁵ Vargas, M. (2008), "Moral Influence, Moral Responsibility", p. 105.

as much a problem as it is a concern), rather, comes from the fact this answer relies on the modularity of Vargas' account : whatever we do with the atypical cases that require explicit deliberation will fall within the purview of whatever theory of normative ethics we adopt. This is of some concern because of the role moral considerations play in the general framework.

For Vargas' account to hold, we need to have a specific kind of agency, one that is responsive to moral considerations. Moreover, we use a theory of normative ethics to settle cases that are complex enough to demand something more than the application of a general rule that has been internalized *and* this theory provides the type and content of the moral considerations we are supposed to be sensitive to. The concern here is that it may be so that whatever the content of the moral considerations, the process of internalization will remain unchanged, but this is an empirical question to which moral psychology must provide an answer. Moreover, it is unclear what exactly those moral considerations are. I do not intend to alleviate this concern here, but an answer would probably have to show that the internalization process really does not depend on the content of moral considerations.

This is, in reality, much the same worry as was expressed in 3.1. The possibility of developing independent normative and descriptive accounts makes revisionism attractive, and it is, with the sidestepping of the metaphysical question, what enables (moderate) revisionism to move beyond the traditional debates, at least at the meta-theoretical level. However, once we examine a substantive, openly revisionist account of the justification of our responsibility practices, it becomes clear that separating descriptive and normative accounts is not without problems, because, at least in the case of the theory examined here, the normative and descriptive are strongly interrelated and dependent on one another. I believe this highlights the main problem of revisionism as conceived by Vargas, which is that while it "looks good on paper", once implemented it becomes clear that the demands we have for a theory of responsibility make the separation of the projects considerably more difficult than we might think at first.

This does not mean that revisionism should be abandoned. Revisionist approaches have been cropping up in the literature for quite some time and are continuing to do so. P. F. “Strawson’s Freedom and Resentment”⁴⁶, which defends the idea that our practices are justified because of our reactive attitudes (emotional reactions to a certain type of action) is one such example, and experimental philosophy⁴⁷ has been doing promising work. Especially in recasting our “folk intuitions”, upon which so much of the work on responsibility and free will rely, in a more naturalistic framework, a project shared by the type of revisionism Vargas advocates. The increasingly naturalistic bent of some work on the question of responsibility promises to clarify many aspects of our moral psychology and thus help us build better descriptive accounts that will appropriately constrain and guide normative accounts. While I believe that such an influence of the descriptive on the normative is to some extent unavoidable, it seems it might raise the concern of the naturalistic fallacy, but as long as we keep this influence restricted to the plausibility constraint outlined in section 1.2, such a concern is unwarranted.

Conclusion

The aim of this paper has been to assess whether an instantiated revisionist approach to moral responsibility⁴⁸ really does track the meta-theoretical account Vargas develops in his 2005 paper. The motivation for this is that moderate revisionism, at the meta-theoretical level, seems to provide a way to move beyond the traditional debates about responsibility and free will *and* to carve out an important place for naturalistic approaches and considerations. This is, to be sure, an attractive prospect, and hence the motivation for assessing whether an actual implementation of revisionism about

⁴⁶ Strawson, P. F. (1960), “Freedom and Resentment”.

⁴⁷ See for instance Nichols who elaborates a naturalist defence of reactive attitudes: Nichols, S. (2007), “After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes”.

⁴⁸ In this case the one developed in: Vargas, M. (2008), “Moral Influence, Moral Responsibility”.

moral responsibility does follow the same structure that lends revisionism its appeal.

However, it seems that while Vargas puts forward a coherent and interesting theory of the justification of our responsibility-characteristic practices, this account, which corresponds to (an aspect of) the normative project a theory of responsibility typically develops, it is far from being independent from the descriptive component. Of course, Vargas never says these two projects have to be completely independent of one another, but he is clear on the fact the normative is constrained by the descriptive, and it would be difficult to argue against such a relationship : whatever we propose, at the normative level, it has to be, at the very least, consistent with what we *can* achieve. The problem I have identified is that it is unclear in which direction the constraint goes.

I do not believe, however, this is a crippling problem for revisionism. For one thing, what the normative account demands of the descriptive, at least in the case discussed here, is not, *prima facie*, completely implausible. The idea that we make responsibility judgements on the basis of internalized norms and that, in general, the responsibility practices are justified because they facilitate and promote the internalization of responsibility norms is plausible, though it really becomes an empirical question. Perhaps this is the main merit of revisionism. If we come to consider theories about responsibility as really being *practical* theories and that to determine whether a given theory of responsibility is adequate what we need is empirical evidence, the debates about responsibility could be usefully reframed.

As an example of what I mean by this, note that the plausibility of Vargas' account depends on empirical considerations about human cognition. In his 2008 paper, Vargas omits to consider adequate empirical support for his view. I have noted, however, that the internalization process he suggests, as well as the end result of that process (the internalization of general decision rules) is, at least at first glance, consistent with the heuristics and biases literature, especially with Gigerenzer and Todd's work⁴⁹. There remains some

⁴⁹ Gigerenzer, G., P.M. Todd and ABC Research Group (eds.) (2000), *Simple Heuristics that Make us Smart*.

work to be done, especially if work about responsibility is to become more sensitive to empirical considerations.

References

- Broome, M. R., L. Bortolotti, and M. Mameli (2010), "Moral Responsibility and Mental Illness : A Case Study", *Cambridge Quarterly of Healthcare Ethics*, vol. 19, p. 179-187.
- Chisholm, R. (2007), "La liberté humaine et le moi", in Garcia, E. & F. Nef (eds.), *Métaphysique contemporaine*, Paris, Vrin, p. 325-341.
- Dennett, D. C. (1973), "Mechanism and Responsibility", in Dennett, D.C. (1981), *Brainstorms*, Cambridge, MIT Press.
- Dennett, D. C. (1981), *Brainstorms*, Cambridge, MIT Press.
- Dennett, D. C. (2005), *Freedom Erodes*, New York, Viking Books.
- Gigerenzer, G., P.M. Todd and ABC Research Group (eds.) (2000), *Simple Heuristics that Make us Smart*, Oxford, Oxford University Press.
- Harnad, S. (2003), "Can a Machine be Conscious ? How ?", *Journal of Consciousness Studies*, vol. 10, n°4-5, p. 69-75.
- Hyslop, A. (2009), "Other Minds", *Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/entries/other-minds/>>, retrieved April 30th 2011.
- Kahneman, D., P. Slovic, and A. Tversky. (1982), *Judgements Under Uncertainty- Heuristics and Biases*, Cambridge, Cambridge University Press.
- Kurzban, R., J. Tooby and L. Cosmides. (2007), "Can Race be Erased ? Coalitional Computation and Social Categorization", *Proceedings of the National Academy of Sciences*, vol. 98, p. 15387-15392.
- McKenna, M. (2009), "Compatibilism", *Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/entries/compatibilism/>>, retrieved April 29th 2011.
- Nichols, S. (2007), "After Incompatibilism : A Naturalistic Defense of the Reactive Attitudes", *Philosophical Perspectives* vol. 21, p. 405-428.
- Strawson, P. F. (1960), "Freedom and Resentment", *Proceedings of the British Academy*, vol. 48.

Implementing revisionism :
Assessing a revisionist theory of moral responsibility

- Trakakis, N. And D. Cohen. (2008), *Essays on Free Will and Moral Responsibility*, Cambridge, Cambridge Scholars Publishing.
- Vargas, M. (2005), “The Revisionist’s Guide to Responsibility”, *Philosophical Studies*, vol. 125, p. 399-429.
- Vargas, M. (2008), “Moral Influence, Moral Responsibility”, in Trakakis, N. and D. Cohen (eds.), *Essays on Free Will and Moral Responsibility*, Cambridge, Cambridge Scholars Publishing.
- Vargas, M. (manuscript), “Building a Better Beast”, http://usf.usfca.edu/fac_staff/mrvargas/Papers/Beast.pdf, retrieved April 21th 2011.