

3

Crossing a Mesh Theory with a Reasons-Responsive Theory: Unholy Spawn of an Impending Apocalypse or Love Child of a New Dawn?

Michael McKenna and Chad Van Schoelandt

In this paper we explore the prospects of crossing a mesh theory of freedom with a reasons-responsive theory. While neither of us is yet prepared to commit to a hybrid view, we find it highly suggestive insofar as each theory might profit from drawing upon the other. In particular, a mesh theory may help to fill an apparent gap in the resources provided by a reasons-responsive theory. At the same time, a reasons-responsive theory may also help to identify more elegantly than a mesh theory alone can the proper extension of free acts. Because the central points we wish to make here apply at a fairly permissive level of generality, we will not develop in much detail either the mesh or the reasons-responsive elements of the hybrid theory to come. A rough characterization of each will serve present purposes. Readers are invited to plug in whichever version of a mesh theory they find most promising (e.g., Frankfurt 1971; Watson 1975; or Bratman 2004a), as well as whichever version of a reasons-responsive theory they find most promising (e.g., Brink and Nelkin 2013; Fischer and Ravizza 1998; Haji 1998b; McKenna 2013; or Sartorio forthcoming2). For ease of exposition, we will simply make do with the most basic features of Frankfurt's original mesh theory (1971), and as well the most basic features of Fischer and Ravizza's reasons-responsive theory (1998). These are, respectively, the most well-known and influential versions of each approach.

1 Two theories of freedom: one inwardly directed, another outwardly directed

Mesh theories account for freedom in terms of a well-functioning harmony between different psychic subsystems leading to action. Exercises of agency are free when they issue non-deviantly from a suitably operating mesh between different elements in an agent's overall mental economy. On Frankfurt's (1971) well-known hierarchical version of a mesh theory, for instance, an agent acts of her own free will when the first-order desire issuing in her action is one with which she identifies and, at a high-order, desires to be effective in leading her to action. Here the mesh is between different orders of desire. When all goes well, there is a smooth transition from characteristic proximal antecedents of action, such as deliberation or intention, to free action itself. But when the mesh is in disharmony, so that an agent's actions issue from desires with which she does not identify and she does not, from a high-order, desire to be effective, she does not act of her own free will. Here, her lack of freedom is accounted for in terms of her psychic elements operating as an upshot of dysfunction.

One observation we wish to make here, which will figure centrally in subsequent discussion, is that mesh theories account for the proximal causal antecedent of free action in terms of internal states of the agent. While of course such a theory will have to allow that an agent stands in the right relation to her environment, the significant dimension of her freedom as highlighted by the theory is accounted for in terms of relations internal to the agent's own psychology. It is in this sense 'inwardly' directed.

Reasons-responsive theories account for freedom in terms of an agent's responsiveness to a sufficiently complex pattern of reasons. An agent acts freely when her action issues from agential causes that are sensitive in this way. Such an approach is a natural extension of the ancient thought that man is a rational animal. On Fischer and Ravizza's (1998) well-known mechanism-based version of a reasons-responsive theory, for instance, an agent acts freely when she acts from her own mechanism of action that is moderately reasons-responsive.¹ Moderate reasons-responsiveness is then explained in terms of an agent's being able to recognize and then react to a relatively complex range of reasons, including moral ones. This is meant to model the exercise of sane, morally competent agency. On such a view, a free agent is one who is 'alive' to reasons to act other than as she does, and she is able to guide her conduct in light of that range of potential reasons. An agent fails to

act freely when she acts from resources that are not alive to a sufficiently rich pattern of reasons.

Another observation we wish to make here, which will also figure centrally in subsequent discussion, is that reasons-responsive theories account for the proximal causal antecedents of free action in terms how an agent is externally related to her environment. Variations in an agent's environment will give rise to different reasons for action. While of course such a theory will have to allow for certain internally related features of an agent's action-generating psychic stew, the significant dimension of her freedom as highlighted by the theory is accounted for in terms of an agent's mode of interacting with elements of the world around her. A reasons-responsive theory is in this sense 'outwardly' directed.

We will soon turn: first to apparent shortcomings with mesh theories, and then to apparent shortcomings with reasons-responsive theories. But before we proceed, we pause to make three important qualifications.

First, we restrict our attention to theories of freedom in our discussion, as we have in our opening remarks. But we mean our treatment to apply to the broader arena of moral responsibility. It is just that, as we understand them, both mesh theories and reasons-responsive theories are best understood as theories meant to account for the freedom or control condition(s) for moral responsibility. Hence, we will attend to these theories of freedom as candidates for explaining the strongest freedom or control condition necessary for moral responsibility. We also note that we are interested in just the sense of moral responsibility bearing upon an agent's accountability for her conduct, which renders her liable to others holding her to account by way of expressions of the reactive attitudes, which can involve the prospect of sanctions and the imposition of obligations to rectify harm done.

Second, it is common to associate both mesh theories and reasons-responsive theories with compatibilism. This is due to the fact that the best-known version of each theory – Frankfurt's and Fischer and Ravizza's respectively – has been advanced in the service of a compatibilist agenda. But it is open to incompatibilists to embrace either a mesh theory or a reasons-responsive theory. They can simply argue that a further condition of acting freely is that an agent's action is not causally determined. Indeed, the literature bears this out.² Thus, we intend that the hybrid theory we shall advance in what follows will be one that is neutral as between compatibilists and incompatibilists.

Third, we acknowledge at the outset that mesh theories and reasons-responsive theories appear to be competitor views at odds with one

another. So in fashioning a hybrid view, we have our work cut out for us. As a matter of historical record that is how they have been received within the philosophical literature – as opposed views. One well-known example is Susan’s Wolf’s (1990) rejection of Frankfurt’s theory, which she characterized as a Real Self theory, in favor of a theory of freedom that involves the ability to act in accordance with the True and the Good. If we take Wolf’s sense of ability to be essentially a form of reasons-responsiveness (to act alive to reasons bearing on the True and the Good), we have a clear case of discarding a mesh theory for a reasons-responsive one. There are, moreover, cases in which the theories apparently yield competing verdicts. For example, consider Frankfurt’s (1971) familiar case of the willing addict, who takes the drug to which he is addicted of his own free will. Frankfurt regards the willing addict as acting freely in the relevant sense, since the effective desire leading him to action is the one that he identifies with and that he wants to be effective. But on a reasons-responsive theory like Fischer and Ravizza’s, the willing addict would *not* act freely (in the sense relevant to moral responsibility) since he would not act from resources that were reasons-responsive.³ Given that his first-order desire leading to action is an over-powering one, he would take the drug regardless of a considerable range of good reasons not to do so.

2 Apparent shortcomings with mesh theories

We turn to apparent shortcomings with mesh theories. We describe them as ‘apparent’ since our intention is not to offer a full defense of the problems we wish to raise for mesh theories. Mesh theorists are likely to have thoughtful ways of resisting the problems we will consider. Our goal is simply to offer a sketch of some difficulties that mesh theories seem to face and then show that a hybrid theory is able to elegantly account for these (perhaps merely apparent) problems. We will consider three problems for mesh theories.

First, we will return to our discussion in the previous section of the opposing judgments about the willing addict issuing from Frankfurt’s mesh theory, in contrast to a reasons-responsive theory. There we were interested in illustrating that the theories yielded different verdicts. That aside, we take it that, on its face, Frankfurt’s treatment of cases like the willing addict counts *against* his theory, and it would count against any mesh theory that would treat similar cases as involving free action, in the sense relevant to moral responsibility. The problem does not seem in any way unique to Frankfurt’s proposal. If the basis for accounting for freedom is action flowing from a harmoniously aligned mesh, then it is

possible for different elements of a proposed mesh to be aligned, even when the causal story is arranged so that the agent who acts is causally driven by ingredients that render her conduct inflexible, in the way the willing addict's drug taking does.

One might suppose that a mesh theorist could push back by distinguishing cases where the action issuing from the mesh is freedom-conferring from cases where the action issuing from the mesh is not freedom-conferring. It might be argued, for instance, that the willing addict is not able to act otherwise when she acts from her harmoniously aligned mesh, and so she does not act of her own free will. Whereas a willing *non-addict* who also takes the drug from a desire with which she identifies and wants to be effective *would* act of her own free will insofar as she is able to act otherwise and freely chooses not to. Fair enough. But notice that to get the desired result for the mesh theory, an adherent has to reach beyond the resources of a mesh theory and appeal to a further factor: the ability to do otherwise. We concur with this general strategy,⁴ but naturally this is because it involves some sort of hybrid approach.

Second, consider cases that are mirror images of the previous difficulty: cases in which an agent acts freely but *not* from a harmonious mesh. Consider a non-addict who takes a drug freely but who does not identify with her desire to take it. Instead, she identifies with her desire not to take the drug. And she wants, at a higher order, her desire not to take the drug to be effective in leading her to action. Moreover, she judges it best that she not take the drug. But she freely takes it anyway. Her case is just a simple case of akratic action.⁵ She acts from weakness of the will. Regardless, she acts freely in taking the drug, and there is no reason to think, if she does act freely, that she is not morally responsible for doing so.

Setting aside the first problem, mesh theorists have a couple of options for attempting to show that these cases are not threats to their theory. One is to argue that mesh theories only promise sufficient conditions for freedom, not necessary conditions as well. Indeed, mesh theorists discharging their arguments in the service of defending compatibilism might retort that all they wish to show is that there exist free acts whose conditions can be satisfied even under the assumption of determinism. The conditions of their preferred mesh theory can all be met even if determinism is true. It might well be that there are *other* free acts that cannot be explained from the resources of a mesh theory, but that is beside the point as regards the metaphysical dispute.⁶ Why? If it can be shown that there are *any* free acts (in the sense pertaining to free will) at a world in which determinism is true, then compatibilism is established.⁷

Mesh theorists, qua compatibilists, who would wish to reply in this way could not be faulted so long as their aim was limited to addressing the metaphysical debate regarding the threat of determinism. But, by proceeding in this fashion, mesh theorists would also leave unexplained a considerable swath of intuitively free action. And this gives rise to the worry that the theory they advance does not properly capture the freedom that, it is assumed, competent morally responsible agents possess. So they would achieve one agenda at the expense of being able to offer a credible account of the actual moral psychology at work in creatures such as ourselves. Furthermore, the problem seems to be an obvious structural problem with any pure (rather than hybrid) mesh theory, and not one that is isolated to Frankfurt's theory alone. Isn't it metaphysically possible for an agent to act freely from an unharmonious mesh, however the mesh is characterized? More forcefully, setting aside arguments for skepticism about freedom and responsibility, isn't it also likely that there are in fact *actual* cases in which free agents really do this? Isn't this just what happens in typical cases of weakness of will, including cases of calmly letting go and giving into temptation, and as well in some cases of recklessness?⁸

Third, consider the curious being Frankfurt calls a *wanton*. A wanton, as Frankfurt characterizes him, is not a person – is less than a person. This is because he does not care about his will (1969, as appearing in Watson, ed.: 327). That is, he has no higher-order desire that any of his first-order desires be effective in leading him to action. Frankfurt calls these distinctive sorts of higher-order desire *volitions*. It is lacking any *volition*, according to Frankfurt, that precludes a being from being a person. Frankfurt grants that it is possible that some adult humans may indeed be wantons. To this, he adds:

In any case, adult humans may be more or less wanton, in response to first-order desires concerning which they have no volitions of the second order, more or less frequently. (328)

Also, in clarifying the nature of a full-on wanton, Frankfurt discusses the case of the wanton addict who has no concern for whether or not his addictive desire for a drug wins out. About this wanton, Frankfurt indicates that either he cannot *or* does not care about which of his first-order desires wins out (329).

We note three things about Frankfurt's introduction of the concept of a wanton. In conjunction, they give rise to two puzzles that, to our knowledge, no one has ever acknowledged much less attempted to

resolve. First, one sense, indeed the primary sense, in which Frankfurt deploys the concept is as a *status* term. Some beings on Frankfurt's view have a status that, as wantons, renders them insufficiently developed to be persons at all. In this sense, a being's status as a wanton functions like an exempting rather than an excusing plea with regard to moral responsibility.⁹ Second, by allowing that an adult human may be more or less wanton with respect to something, Frankfurt allows for deployment of the concept as an *adverbial modifier* with respect to a domain of activity. In doing so, 'wanton' identifies the conduct of one who *is* a person and a morally responsible agent and is just 'wanton' with respect to something she does (or more precisely, wanton with respect to the desires driving what she does). Third, by considering the possibility that a wanton (in the status sense) can, but simply does not, care about which of her first-order desires wins out, Frankfurt allows for the possibility that an adult human might fail to be a person, and so also a morally responsible agent, even if she has the capacity to form volitions and simply fails to do so. Now consider two puzzles.

First, consider the status sense of a wanton. It is easy to understand how it is that a being who is literally incapable of forming any higher-order preferences about how she acts is simply not a candidate for being a person – or at least for being a morally responsible agent. But it is less clear that a being, such as an adult human, who *is* capable of forming such preferences but who simply fails to do so is not a person or a morally responsible agent at all. Why think this rather than that this is a person who is culpable – that is morally responsible and blameworthy – for failing to adopt any attitudes regarding how she cares to act and what sort of person she cares to be? This is puzzling and needs to be explained. While we find this harder to accept, in proposing a hybrid theory we will attempt to explain it and show why it is appealing.

Second, consider the modifier-of-action sense of wanton. Imagine a person who is not only is capable of caring about what moves her to action, but in fact does care about these matters in a considerable sphere of her life. Grant that she is even relatively effective as a morally responsible agent in largely acting in accord with her volitions, and consequently often acts of her own free will in the paradigmatic sense Frankfurt has in mind. But suppose that there is a domain of her life, perhaps involving a simple appetite such as desire for food and drink, in which she is genuinely wanton. She is capable of caring about how and in what ways desires for food and drink lead her to action, but she just does not. Call her Gluttonia. When Gluttonia binge eats or drinks booze recklessly, since she has no high-order preferences about whether

her efficacious desires are efficacious, does she do so of her own free will? Is she morally responsible for doing so? Is she blameworthy? (Let us suppose that there are moral pressures on her not to be reckless with her health in this way.) Frankfurt offers us no clear guidance about such a case. His discussion of wantonness and the machinery he deploys in accounting for freedom and responsibility suggest that on his view Gluttonia does not act of her own free will and so is not morally responsible. But why? She has no will – no effective first-order desire – which she identifies as *hers*, and so she does not act of her *own* free will. And yet, she is unlike Frankfurt's unwilling addict who prefers that her desire not to take the drug be effective and who dissociates with the desire that is effective. This person – the unwilling addict – is on Frankfurt's view clearly unfree in acting as she does. So, maybe Gluttonia is less unfree than the unwilling addict? It is hard to say. Regardless of what Frankfurt is committed to, on this point we believe that Frankfurt's mesh theory needs amending.¹⁰ In our estimation, Gluttonia clearly acts freely and is morally responsible, although of course it is less than an ideally free exercise of her free agency. Her behavior involves a kind of negligence in which she fails to form proper attitudes about herself and her motivations that as a competent morally responsible adult person she should form.

Before proceeding, we wish to emphasize that in our estimation the general shape of the problems we have identified for Frankfurt's theory are not unique to his version of a mesh theory. They are liable to infect any pure version of a mesh theory. The first two problems are fairly easy to apply to other mesh theories. The first has to do with compromises to freedom when the mesh is in place. So long as the causal story generating the mesh is open to cases in which an agent is 'straight-jacketed' into operating from that mesh, as she would be were her conduct driven by severe addiction, then a mesh theory is open to the basic challenge we have identified here. So too for the second problem. If an agent can freely act from resources that fail to mesh harmoniously – say, in circumstances in which the agent retains the ability or general capacity to act from a harmonious mesh – then the theory will lack the resources to account for free action in these cases. This will be so unless it builds upon resources external to the elements that make it a mesh theory, like the ability to do otherwise.

Making a case for generalizing the third problem to other mesh theories is a bit more challenging. Space will not permit a proper defense of our claim here, but we would note that one thing that makes any mesh theory attractive is its means of identifying a distinctive feature of

persons when they function as morally responsible agents. The psychic elements of the theory offer some way of capturing the complex inner life of a person. Frankfurt does it in terms of the relation between orders of desire, and how as reflective creatures, we are able to adopt a critical attitude toward and to care about our own motivations. Watson (1975) does it in terms of different sources of motivation, some arising from our motivational system and animated by, among other things, our basic appetites; some arising from our valuational system, whereby we are able to value some courses of action rather than others. Bratman (2003) does it in terms of guiding intentions or plans in relation to which others either are or are not, can or cannot be integrated. In any of these cases, we will be able to identify a status sense of a wanton, either as one incapable of the integration pertinent to the proposed mesh, or simply as one who never exercises that integrative capacity. We will also be able to identify a modifier sense of a morally responsible agent's acting in a wanton fashion by simply failing to align her different elements within her psychic structure out of mere negligence when she would be perfectly capable of doing so. Gluttonia, for instance, was characterized as one who was able to form higher-order preferences about her gluttonous motivations but neglected to do so. She could instead be characterized as one who is able to form and commit to values about her gustatory cravings that would thereby come into conflict with her more base motivations for consumption. Or she could instead be one who is able to form guiding intentions about health and moderation that would result in conflicts with her various plans to binge eat and drink.

3 Apparent shortcomings with reasons-responsive theories

We turn now to challenges for reasons-responsive theories. Here too we wish to begin with the same disclaimer. We describe the problems we shall identify for reasons-responsive theories as apparent shortcomings, since our intention is not to offer a full defense of the problems we wish to raise. Reasons-responsive theorists are likely to have thoughtful ways of resisting the problems we will consider in this section. Our goal is simply to offer a sketch of some rather obvious difficulties that reasons-responsive theories face and then show that a hybrid theory is able to offer an elegant way of accounting for the (perhaps merely apparent) problems we identify. We will restrict our attention to two potential problems.

First, consider Frankfurt's introduction of the wanton. What makes it so easy for Frankfurt to be able to identify them? The wanton is displayed

as a conceptual possibility from the resources of Frankfurt's mesh theory simply because the structural features of a mesh help to bring into relief that there could be beings whose internal life is in some way *unlike* what is distinctive about the inner life of creatures like ourselves; unlike us, they might lack the requisite structure altogether. And the elements implicated in explaining wantonness – such as not adopting a certain sort of higher-order attitude toward one's motivations – do seem to capture by contrast something important about the life and experience of being a free agent.

What reasons-responsive theories seem to lack are the resources to capture an aspect of ourselves as free agents, and as well how we experience ourselves as free agents. We really do often struggle over the role our own motivations play in issuing in action, and we do come to care about, at a higher-order, the causal efficacy of those motivations. Or, to put it in terms instead amenable to Watson's (1975) mesh theory, we really do sometimes find ourselves with an internal struggle over sources of motivation arising from our nature as valuing beings as in tension with sources of motivation arising from our wide array of desires and (often base) appetites. Mesh theories get right that we are like, and they get right that our freedom can be 'problematized' as a result of how these elements might or might not be aligned.¹¹

As persons who are free and morally responsible agents, our internal psychic architecture is complex. It encompasses inner elements sometimes working in harmony and supporting one another, and at other times working in disharmony and in opposition. In this distinctive way mesh theories accurately and vividly capture something important about how we experience our agency as free. And phenomenology aside, they also identify an especially salient feature of how our free agency does often internally function. Sometimes pertinent psychic ingredients figuring in a mesh are proximal casual antecedents to (putatively) free acts. Contrary to what Frankfurt contends, when we act of our own free will, we might not *always* act freely in accord with the desires we would prefer to act on. But surely *sometimes* we do, and when we do there is a sense that it is a more desirable expression of ourselves and our freedom. By contrast, reasons-responsive theories work just with agents and the spectrum of reasons that are pertinent to free action.¹² As such, they offer no natural way to carve at the joints, so to speak – the major architectural elements comprising the internal psychic elements of free persons of the sort we take ourselves to be.

Second, consider the notion of self-creation. Here we do not mean what Galen Strawson (1986) and some others have suggested is a

requirement of moral responsibility – that one literally be able to be a (complete) cause of oneself. All we are interested in is the familiar thought that many persons purposefully invest in shaping themselves. Consider, for instance, the remark that so and so is self-made. Robert Kane (1996) contends that the most fundamental kind of free action involves self-forming-acts (SFAs). In these cases, Kane contends, not only do we freely settle what to do in conditions of uncertainty about such matters as what we find most valuable, we also contribute to the shaping of our very selves.¹³

We find Kane's notion of self-creation eminently credible. Something like Kanean SFAs identify distinctive exercises of our free agency (regardless of whether they require libertarian satisfaction conditions as Kane would contend they do). Here too, mesh theories, in contrast with reasons-responsive theories, seem to provide resources helping to identify challenges free agents often face when performing SFAs. Kane makes use of a now famous example of a businesswoman making her way to an important meeting when she comes upon a person being assaulted. She must choose between, and in the light of, reasons and values of self-interest and reasons and values of morality. In doing so, Kane argues, if this is genuinely an SFA, she also contributes to the shaping of herself. Now, in the case Kane describes, we can imagine a scenario in which all that was on the woman's mind was whether to help or continue on her way. She might even have consciously thought in terms of what was most valuable. If that is how she engaged in the situation, a reasons-responsive theory is at least on equal footing with a mesh theory in accounting for the woman's freedom and responsibility. But suppose instead that we imagine the case to go like this: The woman actually considers whether she wants to be the kind of person who would abandon a person in need for personal gain, or instead whether her altruistic feelings for others will always hamper her from success in her own life. Suppose also that in doing so she acknowledges her own motivations and fears as having a grip on her: she finds herself in another struggle regarding how she would prefer her own motivations to operate in relation to, say, what she most values. If this is the context of her struggle in this SFA, a mesh-theory helps to identify how, in a process of self-formation, a person might think and act in light of her own sense of her internal psychic architecture. Reasons-responsive theories offer no special resources to help us identify or account for such exercises of free agency: exercises that are precisely *about* freely acting upon one's own internal psychic structure. These special cases of self-correction and self-learning are readily captured

from the resources of a mesh theory, but not from a reasons-responsive theory.

There is an all-too-easy response that reasons-responsive theorists might offer in reply to the two problems we have raised here. They might reply that a reasons-responsive theory has no prejudices about the spectrum of reasons to which a free and morally responsible agent can be responsive. If free persons have the sort of complex internal psychic structure that mesh theories identify, then those elements will have a role to play in the sorts of reasons to which a person can be responsive. The structure that, it is alleged, is phenomenologically salient and causally operative in the exercise of many free actions will be revealed in the kinds of reasons to which an agent can be responsive. If, for instance, a person struggles with her powerful desire for food, this will be reason-providing for her in so far as certain courses of action will either enhance or instead diminish those desires. As for the issue of self-formation, consider Kane's businesswoman. If one of the considerations to which she is alive concerns whether she wants to be the sort of person who gives in to her altruistic feelings, then a reasons-responsive theory can include, among the reasons to which an agent might be responsive, reasons bearing on this question and on action aimed at making herself come to be (or not) a certain way.

Regrettably, for the reasons-responsive theorist, the preceding reply really is too easy. Our imagined reasons-responsive theorist is probably correct that if agents have the sort of complex psychic structure mesh theories allege that they have, this will show itself in the spectrum of reasons to which free and responsible agents are responsive.¹⁴ But a good theory of human agency that accounts for freedom and responsibility should not just capture the extension of exercises of free agency. It should capture them in explanatorily illuminating ways, and it should identify relevant structure, if it really is there and really does have a causal role to play in how free actions are generated. By failing to attend explicitly to the internal operation of the complex structure of the agency of persons such as ourselves, pure reasons-responsive theories seem to fall short of capturing something that needs to be captured.

4 Introducing a hybrid view

We will now introduce a hybrid theory of freedom by drawing upon the advantages of both mesh theories and reasons-responsive theories. By doing so, we hope to show how the apparent problems facing each approach can easily be addressed by drawing upon the other. We

believe that our proposal is not mere patchwork philosophy, simply pasting together in *ad hoc* fashion two different theories. As a pre-theoretical observation about what free and responsible agency is like, a hybrid theory seems to capture its nature better than either approach can on its own. Free agency, we contend, has two faces – inner and outer. Exercises of free agency *do* seem to issue from, at least often if not always, a process with an inner complexity involving the inter-animation between different psychic subsystems, preferably working in harmony rather than not. But the exercise of free agency *also* seems to issue from a process that situates an agent more or less reliably in her external environment. This allows her, qua practical agent, to be sensitive to how reasons for action might vary.

There is, we contend, a natural and illuminating way of fitting the two approaches together into a comprehensive theory of free agency. Here is our proposal, to a first approximation:

A person acts freely in the strongest sense necessary for moral responsibility only if 1) she possesses the ability to act from a suitably integrated and harmoniously functioning psychic mesh; 2) in acting as she does she is appropriately reasons-responsive; and 3) the reasons-responsive resources from which she acts must be accessible to consideration from within the framework of the agent's mesh.

Call this proposal *Hybrid-Mesh-RR (HMRR)*. Much needs clarifying in this proposal, which we cannot work out in this paper. Here we simply note the various bits that need attention. To begin, the qualification 'suitably integrated' needs unpacking. Different mesh theories will do so in different ways. Frankfurt would do so in terms of identification and higher-order desires. Likewise, the 'appropriately reasons-responsive' needs unpacking. Different reasons-responsive theorists will do so in different ways. Fischer and Ravizza would do so in terms of their notion of moderate reasons-responsiveness. There is as well the notion of accessibility in the expression 'accessible to consideration from'. What is the sense in which some features of an agent's psychology are available or accessible to her? (Cases of self-deception, for instance, could prove difficult to accommodate.) There is also a further worry about what 'consideration' amounts to in the expression 'consideration from within the framework of the agent's mesh'. On a view like Frankfurt's, an agent's considering the reasons-responsive resources from which she acts would be a matter of her psychic resources allowing her to form a pro or con higher-order preference about those reasons-responsive

resources. Finally, there is the sense of ability enlisted in the requirement that an agent possess the ability to act from a suitably integrated and harmoniously functioning mesh. This might be the most controversial ingredient in our proposal. Appeal to ability in any account of freedom gives rise to familiar metaphysical challenges. Does it merely pick out a general capacity, or a more narrowly specified power involving the unimpeded opportunity to be exercised in the context in which one acts? Compatibilists might be more inclined to embrace the former and incompatibilists the latter. We leave this detail along with the others mentioned here unsettled.

HMRR is really just a first pass at a theory of freedom. In truth, it is better thought of as a skeleton of what a hybrid theory of freedom might come to rather than a proper theory in its own right. Nevertheless, in its nascent form, it still is informative enough to shed light on how to deploy it to resolve the difficulties arising in the last two sections for mesh and then reasons-responsive theories respectively. Readers will recall that in our introductory remarks we claimed that the resources of a reasons-responsive theory might be used to aid a mesh theory in capturing more effectively the proper extension of free acts. At the same time, we claimed, the resources of a mesh theory might be used to aid a reasons-responsive theory in filling a lacuna in their overall account of free agency. By briefly retracing the path we took in the last two sections, we believe we can easily establish these claims.

Consider the first two apparent problems we identified for mesh theories. In some cases an agent acting from a harmonious mesh, such as a willing addict, seems unfree when a mesh theory would yield the result that the agent does act freely. In other cases, such as cases of weakness of will, an agent does not act from a harmonious mesh, and yet it seems that she does act freely. It appears that a mesh theory cannot on its own account for an agent's freedom in such cases. But reasons-responsive theories are tailor-made to capture these sorts of cases. The willing addict is not free according to a reasons-responsive theory because the proximal casual sources of the agent's acting as she does are not sufficiently sensitive to reason. The willing addict would persist in taking the drug to which she is addicted, regardless of whether any of a wide swath of candidate reasons to do otherwise were applicable to her context of action. Consider now the weak-willed non-addict who does not act from a harmonious mesh, in comparison with the willing addict who does. The weak-willed non-addict, the reasons-responsive theorist can argue, is free while the willing addict is not simply because there are a range of reasons that would move the former but not the latter to refrain from

taking the drug (McKenna 2013). Her freedom is displayed by her greater flexibility in responding to reasons, in comparison to an addict.

But why regard the preceding treatment as grounds for supplementing a mesh theory with a reasons-responsive theory? Doesn't the latter just supplant the former? After all, the reasons-responsiveness is doing all the work in identifying the correct extension of free action. True. But what a mesh theory offers that a reasons-responsive theory apparently lacks is a natural way of identifying the psychic predicament of the willing addict and the weak willed non-addict. Both in different ways experience their agency as compromised by virtue of elements of their own inner lives. And both do act from proximal causes that involve the influence of the pertinent sort of psychic structure.

Now consider the person who is wanton with respect to some domain of activity, like Gluttonia. Gluttonia has no preferences about what moves her to action in consuming food and drink as she does. On this score, we argued that Frankfurt's thesis was incomplete in that it did not offer guidance as to whether we should regard Gluttonia as acting of her own free will or not. Moreover, we claimed that the correct verdict should be that Gluttonia does act freely and is morally responsible, so long as she is able to adopt a preference regarding her motivations for action but fails to do so. Negligence on her part is in no way freedom-defeating or more generally exculpatory with respect to her responsibility. Our proposed HMRR helps to accommodate our judgment about Gluttonia. We can assume that she has the ability to act from a suitably integrated mesh. Indeed, we stipulated that in other arenas of her life she actually does so, and we may specify that there is nothing but negligence in the way of forming the relevant structural pieces to cease her wanton ways. Moreover, we can grant that when she binge-eats and drinks she is reasons-responsive. If so, then so long as she can avail herself of a stance whereby she would consider the desirability of her being so moved, there is no reason to think her freedom is expunged simply because she neglects to do so.

Here too we contend that it is the addition of a reasons-responsive condition that gets the extension of free agency correct. It is what enables a mesh theorist to get the verdict about an agent like Gluttonia that she does act freely. Her being reasons-responsive establishes that she possesses and acts from a significant form of agential control. Her negligence alone does not preclude her acting of her own free will so long as she is able to adopt a critical stance toward her own sources of motivation. One might even further amend the hybrid theory by pointing out that an agent's having access to considerations whereby

she could evaluate her own motivational states is *itself* reason-giving. And so an agent like Gluttonia can also be more or less reasons-responsive with respect to whether she actively considers or reflects upon her own motivational states in higher-order Frankfurtian fashion. If, on the other hand, an agent is *not* sufficiently reasons-responsive in this domain, this would suggest that she is after all screened off from being able to consider her own motivational resources. This might be grounds for thinking that she does not act freely.

What of the difficulties we identified for reasons-responsive theories? Our hybrid proposal requires that a free agent be *capable* of adopting and acting from a suitably integrated mesh. With respect to any actions she performs, there is no requirement that she *must* act from a harmoniously operating mesh. But if an agent has the resources to do so, we can account for how it is that an agent might experience her freedom or lack thereof as arising from the functional or instead dysfunctional interplay between elements of her own psychic resources. The account can explain, for instance, why when an agent not only does not but cannot act from an integrated mesh she will experience her action as beyond her control. And we can as well make room for how in fact an agent's actions often are generated from well functioning or instead poorly functioning elements of her internal psychic structure. As for the kind of agency involved in self-formation, agents really do sometimes purposefully aim their efforts at self-improvement or self-correction directly at their own internal psychic forces. Now, an agent might only be able to do so and then choose freely if she is reasons-responsive. Indeed, we think this is so. But among the reasons to which she might be reasons-responsive are reasons regarding her own internal psychic nature.

This leaves only one difficulty unaddressed. It involves a puzzle regarding Frankfurt's treatment of wantons (in the status sense). Frankfurt allows that an adult human being could be less than a person, simply by failing to care about her own motivations that lead her to act as she does. This is possible on Frankfurt's view even if the wanton is able to care in the requisite way but does not. This *does* seem puzzling. But we wish to explore the idea and offer a tentative defense of it, although not quite in the form Frankfurt cast it. Frankfurt makes caring about one's own sources of action (having volitions about them) a requirement for *personhood*. We do not wish to go this far. A less extreme view would be that a person is a *free and morally responsible agent* only if she *does* care about (has volitions with respect to) her own sources of motivation, and so has *actually* adopted a stance regarding her own agency. Just now we expressed this in terms internal to Frankfurt's own theory. But the point

is amenable to other mesh theories. The idea would be that a person is a free and morally responsible agent only if she engages the resources of a mesh in a way that commits her to a point of view on or attitude about her own agency.

Why commit to a view like this? Why not allow for free agents who are so reckless or negligent that they simply never take any steps at all to fashion any sense of themselves? If they can but don't, then why aren't they just free and responsible but despicable? Aren't they just louses to blame for squandering their lives? We certainly feel the pull of this judgment. But here is a reason to commit to a stronger thesis. One appealing condition on autonomous agency according to philosophers like Ishtiyaque Haji (1998b) and Alfred Mele (1995) is that autonomous agents must satisfy an authenticity condition. It is not enough that an autonomous agent simply act in a way that involves some sort of control (maybe cast in terms of reasons-responsiveness). She must do so from within the context of values or principles that she herself has come to take on as her own (even if she acts contrary to them). The worry is roughly that if her own evaluative standpoint is not one she has come to authentically possess but instead has been forced upon her or acquired and then sustained through some deviant means, then when she acts, her agency is not really an expression of any point of view that *she* had a hand in fashioning. She is, in a sense, a kind of puppet or little robot pushed around by an orientation that was out of the reach of her own agency, even if that orientation is now hers.

We do not mean to suggest that a hybrid theory should adopt the same kind of proposal either Haji or Mele develop (in different ways). We only wish to point out an incentive for requiring of a free agent that she have actually adopted, through a suitable process (or lack of an objectionable process), some point of view or set of concerns or commitments regarding her own agency. The incentive is similar to the one motivating Haji or Mele. Without an agent's actually having done this, there is in a certain respect no standpoint on herself or an internal psychological orientation that is her own. And this renders her a kind of wanton. In this sense, there is no one to identify as *freely* acting in accord or out of accord with how she wishes herself to be, since there is nothing she wishes to be or wishes not to be. This finding is similar to one Fischer and Ravizza (1998) have adopted in their theory of (the freedom condition for) moral responsibility. On their view, morally responsible agency has a subjectivity requirement in that a person is not a morally responsible agent unless she comes to see herself as such. The subjectivity requirement we are proposing here in fleshing out

Frankfurt's claim is a different one. Regardless, it is a subjectivity requirement, and it is a compelling one. It is hard to imagine a concrete case of an actual human adult, one who is not severely impaired in some way, who would be able to take up some stance toward her own agency but simply neglects to do so *altogether*. Perhaps this being, whether an actual person or not, would be an extreme sort of buffoon or instead a moral monster. In any event, she is a barely imaginable possibility, and even less imaginable as free and responsible.

5 Moral identity and the real self

Our hybrid proposal has a further virtue we have yet to mention. Mesh theorists have drawn a distinction between actions flowing from an agent's real self, to use Wolf's (1990) language, and actions that are in some manner alien to the agent. The former are regarded as freedom-and-responsibility-conferring; the latter are regarded as freedom-and-responsibility-defeating. Frankfurt, for instance, writes of the unwilling addict's making his desire not to take a drug 'more truly his own' while his desire for the drug, which moves him to action, is a 'force other than his own' (329). But theorizing in this manner leads to implausible results. Addicted agents, willing or not, act from their *own* desires, as do weak-willed non-addicts. Put more simply, *they* are the ones who act, and from desires that are *theirs*, regardless of whether or not in doing so they act freely. It stretches what needs to be said too far to contend that there is some sense in which it is not really the agent who is acting when she acts from an unwanted compulsion or addiction. Nor is it any more credible to contend that it is the agent who is acting but from a desire that is not hers. As Mele (1995: 117) has remarked, if it is not the agent's, then whose is it?

What a mesh theorist gets right is that an agent who simply has no moral identity, like a wanton, cannot plausibly be regarded as acting freely and being morally responsible for what she does.

Nor can an agent who possesses a moral identity be so regarded if instead she acts from resources that are screened off from her moral identity. It is thus credible to contend that a free and morally responsible agent simply must have a point of view on her own agency by way of her *possessing* an integrated mesh – even if it is poorly integrated rather than integrated well. She can thereby have some relation to herself as a practical being in the world. One who is screened off from the sources of her actions so that she cannot recognize or assess them does not act freely. This is so, even if the screened-off elements

are reasons-responsive. A free agent cannot respond to reasons only via the 'monkey on her back', even if that 'monkey' does respond to a significant range of reasons.

On our proposal, if an agent is suitably reasons-responsive, she might still act freely and be morally responsible for how she acts when she acts in ways that result in psychic disharmony, as in cases of weakness of will. What the hybrid view requires is that the resources leading to her action – even moderately reasons-responsive ones – are accessible for assessment from the agent's psychic mesh. This way, as a person with a moral identity, she is able to adopt some attitude or other toward, as Frankfurt might put it, the forces that move her.

6 Conclusion

Both mesh and reasons-responsive theories offer virtues for a comprehensive theory of free agency and moral responsibility. Mesh theories get something correct about the inner nature of free agency and the exercise of it. Reasons-responsive theories get something correct about the outer nature of free agency and its exercise. But both also face apparent problems. We have argued that a hybrid theory can settle the problems rather easily, since the problems arising from a mesh theory can be accounted for with the resources of a reasons-responsive theory. And the problems with a reasons-responsive theory can be accounted for with the resources of a mesh theory. The emerging theory is, moreover, motivated independently by reflecting on how it seems at a pre-theoretical level our free agency really is. It does seem to have two faces, an inner and an outer one. When we act freely, we act from internal resources that can sometimes work in harmony, thereby enabling us, perhaps even enhancing our freedom and our sense of it. At other times these internal resources can function in opposition and thus impede us and compromise our own sense of our efficaciousness. At the same time, we also act in the context of an outer environment in which our reasons for acting are at the mercy of conditions of the world beyond us. We are sometimes well-equipped to respond in ways that manifest our freedom, and yet other times we are impoverished in ways that show our freedom to be diminished.

We have proposed a kind of architecture of free agency meant to fold the preceding considerations into a single account. We leave it to others to settle upon what sort of offspring this turns out to be. Some will no doubt regard it as an unholy spawn. Others will likely regard it as a love child. We have simply offered it up as a proposal worth exploring.

Notes

For helpful comments on this paper, we would like to thank Mark Balaguer, Phoebe Chan, Ish Haji, Terry Horgan, Al Mele, Carlos Moya, Sergi Rosell, Carolina Sartorio, David Shoemaker, Tamler Sommers, Chandra Sripada, and Benjamin Mitchell-Yellin. McKenna presented a version of this paper at NOISE (New Orleans Invitational Seminar in Ethics), arranged by David Shoemaker and sponsored by Tulane University's Murphy Institute. We also profited from the many noisy comments from members of the audience. Michael McKenna would like to thank the Templeton Foundation, and Chad Van Schoelandt would like to thank the Institute for Humane Studies, for support in writing this paper. Finally, we wish to thank the editors of this volume for arranging this collection and inviting us to contribute.

1. We set aside two details crucial to a careful representation of Fischer and Ravizza's view. One is their commitment to a mechanism-based view. This detail matters for the execution of their argument for a reasons-responsive theory, but it need not detain us here. Another is that they do not write in terms of acting freely or acting of one's own free will. Instead, they write in terms of acting with a certain kind of control, guidance control. Again, while this is crucial for their position, it is a detail that we can set aside here.
2. For example, see Kane (1996) and Ginet (2006).
3. We set aside here and in the remainder of this paper the pertinence of *indirectly* free acts. Indirectly free acts are acts that acquire their status as free by way of other *directly* free acts. Indirectly free acts 'trace' their freedom to other prior free acts. An addict might on a view like Fischer and Ravizza's act freely even if she is not reasons-responsive at the time if at an earlier time she freely made herself an addict by purposefully setting out to do so. (The distinction between direct and indirect is more commonly applied to moral responsibility.)
4. But to avoid any misunderstanding, we should note that we do not endorse appealing to an ability to do otherwise. One of us (McKenna 2003) rejects this approach. And, as many readers are likely aware, this is also not an approach Frankfurt would accept in light of his other important work (1969) on this topic.
5. Related cases are raised against mesh theories by John Martin Fischer (2012: 129–30), Ishtiyaque Haji (1998b: 74), and Michael McKenna (2011: 181).
6. One of us (Van Schoelandt, n.d.) has argued that mesh theorists have at least some resources to address the problems we have identified here.
7. Of course, this would be so only if the view advanced could withstand arguments such as the manipulation argument for incompatibilism. (We are thankful to the editors for raising this point.)
8. For an expanded discussion of this problem for mesh theorists, see McKenna (2011). For an effort to save mesh theorists from the problem, see van Schoelandt (n.d.).
9. Exemptions show that for some reason an agent is incapacitated for morally responsible agency and so is not a candidate for being blameworthy or praiseworthy for anything at all.
10. For one proposed set of amendments regarding wantonness, see Van Schoelandt (n.d.).

64 *Michael McKenna and Chad Van Schoelandt*

11. For an insightful discussion of various internal threats to freedom, see Benn (1998: chap. 8).
12. Or instead, as Fischer and Ravizza might put it, with mechanisms owned by agents and the spectrum of reasons pertinent to free action.
13. Cf. Buchanan (1979); Knight (1922: 458–9); Schmitz (1995: chaps. 3–4).
14. This is how one of us (McKenna) had previously thought a reasons-responsive theorist could account for the elements involved in a mesh theory like Frankfurt's.