McGeer Victoria (Orcid ID: 0000-0002-7328-0563)

1
MS: Scaffolding Agency
COVER PAGE

MS ID: EJP-17-361.R1
Title: Scaffolding Agency: a proleptic account of the 'reactive attitudes'

Author: Victoria McGeer
Email: vmcgeer@princeton.edu

Affiliations:
University Center for Human Values,
Princeton University
5 Ivy Lane
Princeton NJ 08540
USA

School of Philosophy,
College of Arts and Sciences
Australian National University
Canberra, ACT 0200
Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ejop.12408

Scaffolding Agency: a proleptic account of the 'reactive attitudes'

Victoria McGeer

ABSTRACT: This paper examines the methodological claim made famous by P.F. Strawson: that we understand what features are required for responsible agency by exploring our attitudes and practices of holding responsible. What is the presumed metaphysical connection between *holding responsible* and *being fit to be held responsible* that makes this claim credible? I propose a non-standard answer to this question, arguing for a view of responsible agency that is neither anti-realist nor straightforwardly realist. It is instead 'constructivist'. On the 'Scaffolding View' I defend, reactive attitudes play an essential role in developing, supporting, and thereby maintaining the capacities that make for responsible agency. While this view has relatively novel implications for a metaphysical understanding of capacities, its chief virtue, in contrast with more standard views, is in providing a plausible defense of why so-called 'responsible' agents genuinely deserve to be treated as such.

KEYWORDS: responsibility; scaffolding; capacities; dispositions; skills; reason-responsiveness; Strawson; reactive attitudes; desert

---

Introduction:

In "Freedom and Resentment' (1962), P.F. Strawson famously defended the following methodological claim: that philosophers can make real progress in understanding what traits or capacities are required to be a responsible agent by focussing on our attitudes and practices of holding responsible. Prima facie, the claim is puzzling and interpreting it has been the source of some controversy in the responsibility literature. What is Strawson's understanding of the connection between our attitudes and practices of holding responsible and actually *being* responsible -- i.e. being fit to be held responsible -- that would make such a claim credible?[1] More to the point, how should we understand this connection if we're to capitalize on Strawson's insight? This is the issue I address in this paper.

Of course, the *negative* message of Strawson's paper is relatively clear. It is that philosophers make no progress in understanding the nature of responsible agency via their

abstract and arid preoccupation with the question of whether such agency is threatened by the metaphysical thesis of determinism. Instead, we should focus more theoretical attention on our everyday attitudes and practices of holding responsible, where the problem of responsible agency has some real and pressing significance for us. Indeed, it has such significance that our attitudes and practices of holding responsible are infused with a range of powerful emotions – emotions that only seem appropriate in relation to the doings of responsible agents. Strawson calls these emotions 'reactive', including amongst them: gratitude, resentment, hurt feelings, indignation, guilt, shame, remorse, forgiveness and certain kinds of love. Thus, he refers to our ordinary attitudes and practices of holding responsible as '***reactive*** attitudes and practices': attitudes and practices infused with this special class of responsibility-sensitive reactive emotions.

But what of the *positive* view defended (or implied) in Strawson's paper? Why should our myriad ways of reacting to responsible agents -- agents we take to be responsible -- shed any light on the nature of responsible agency in itself?

I begin this paper by characterizing, in Section 1, two standard answers to this question Strawsonian thinkers provide, outlining the substantially different ways in which they hope to accommodate Strawson's key insight. In Section 2, I explain why I find each of these views problematic so far as they fail to satisfy, each in their own way, a key condition that Strawson himself imposed on any adequate theory of responsibility -- namely, that it provide an intuitively acceptable way of addressing what he characterized as the "pessimists'" concern with genuine desert (1974, p. 3). In light of this failure, I suggest that Strawsonians must find a new way forward. In Section 3, I begin this constructive work by developing a distinctive skill-based account of the nature of 'intelligent capacities'. Building on this account, I then return in Section 4 to the topic of responsible agency, where I lay out an alternative 'Scaffolding View' of the connection between responsible agency and reactive attitudes. Finally, in Section 5, I conclude with an assessment of how well this views fares in relation to the standard views on the critical problem of desert.

Section 1: Two Standard Strawsonian Views

Though a significant number of philosophers have found inspiration in Strawson's work, leading to a proliferation of so-called 'Strawsonian' accounts of responsibility, I discern within this literature two standard ways of understanding the metaphysical connection between *being* responsible and our attitudes and practices of *holding* responsible. I dub these the 'Conventionalist View' and the 'Indicative View'.[2] Individual philosophers may not conform precisely to either one of these 'standard' views; there is often some wobble between them. But my purpose here is not primarily exegetical; it is rather to clarify and examine the space of conceptual possibilities. The key point, emphasized in my discussion below, is that these distinctive metaphysical views imply quite distinctive answers to the question posed above: namely, why it would be epistemologically fruitful to focus on our attitudes and practices of holding responsible to gain an understanding of the nature of responsible agency itself.

1.1 The Conventionalist View[3]

The central contention of the Conventionalist View is as follows:

> What it is to *be* a responsible agent is simply to have whatever it takes to be ***deemed an 'appropriate' target*** of reactive attitudes and practices, as determined by the generally accepted norms (whatever those happen to be) that govern those attitudes and practices.

This view is metaphysically anti-realist (or stipulative) in the following sense: According to Conventionalists, there simply is no underlying objective responsibility-making feature of agents that our reactive attitudes are presumed to be tracking; hence, there is no objective responsibility-making feature of agents that shapes, anchors or ultimately governs whatever norms emerge in our attitudes and practices of holding responsible. These norms have a life of their own. Of course, they may serve any of a variety of cultural or social functions;

and they may take the shape they do as a matter of historical contingency. But however they are determined, they are not determined by independent responsibility-making features of agents. The explanatory arrow goes the other way. It is the norms themselves that determine or stipulate the features of agents that make them into responsible agents, regardless of how motely such features may be.

Some analogies may help in making this idea clear. Consider the category of things that are 'fashionable'. In any given culture, at any given period of time, there are a wide variety of things and/or doings (clothes, music, paintings, performances, manners, pastimes) that are fashionable. The fashionable is a real property-implicating category – that is to say, it picks out actual properties in the things and doings themselves. If we are appropriately in the know (i.e., understand the norms that govern what's fashionable), then we can certainly sort the various things we encounter into the fashionable and unfashionable, according to properties those things actually have. Happily, we can also change something that is unfashionable into something fashionable simply by altering its (objective) properties. Furthermore, in case this has gone unnoticed, the fashionable is clearly a socially significant category: it is important to us in a variety of ways, important enough that we have a number of emotional responses 'appropriately' tuned into the relevant properties in our environment. (Not for nothing is the unfashionable regarded as 'naff', 'geekish' or 'daggy' – and our feelings towards those so designated, especially amongst certain constituencies, often a combination of pity and humorous contempt.) And yet despite all this, there are no objective features of things in the world that somehow shape or govern what norms to embrace regarding what is fashionable (deluded fashionistas may disagree with this point, but I trust the rest of us would not). The explanatory arrow goes the other way. It is the norms themselves – however these are determined – that determine or stipulate the features of things or doings that (here and now) make them fashionable.

One more analogy: consider the category of sentences that are 'grammatical'. In any given language, at any given period of time, certain sentences are grammatical; others are not.

Again, this is a real category, picking out actual structural features of sentences in and of themselves. So if we understand the norms, or rules, that govern what's grammatical in a language, we can sort the sentences of that language accordingly; likewise we can turn ungrammatical sentences into grammatical ones by adjusting the relevant (objective) properties. The grammatical is also a socially significant category: it is important to us in a variety of ways, important enough that we may even have emotional responses (whether pro or con) 'appropriately' tuned into the relevant features of spoken and written language ('doesn't he sound like a toff, minding his p's and q's?!"). Yet here I trust it is even more obvious that there's nothing about the way words are put together in and of themselves that somehow shape or determine what norms (or rules) to embrace regarding what is grammatical in a given language. Again, the explanatory arrow goes the other way. It is the norms themselves – however these are determined[4] – that determine or stipulate the features of sentences that (here and now) make them grammatical.

To introduce a bit of philosophical jargon, we might say the category relevant properties of things that *make* them fashionable or grammatical are *response-dictated;* more precisely, they are response-dictated in keeping with the generally accepted norms of a shared practice. Likewise, we might say that the 'fashionable' and the 'grammatical' are *response-dictated categories* – i.e. categories of things whose category-determining properties are response-dictated.[5]

Now there are three things worth noting in connection with these response-dictated categories:

(1) Individual fallibility: Individuals can have perfectly good knowledge about the objective properties of things and yet be wrong in their judgements as to whether something has the relevant response-dictated property for category membership. For instance, I may think that an ensemble composed of a tucked-in polyester shirt, bermuda shorts, white knee-socks, and tan hush-puppy shoes will cause my fashion-conscious neighbour to sigh with envy, but I am

wrong! My understanding of the norms governing what's fashionable is just completely off-base.

(2) Collective infallibility: It is *not* the case that *everyone* could be wrong in this way. So, for instance, if my fashion-conscious neighbour happens to be the editor of Vogue magazine, then she and her fashionista colleagues *dictate* what it is to be fashionable around here: they are the norm-setters. Consequently, if my ensemble meets her approval, I need do nothing more to enter the ranks of the fashionable that what I am already doing: lo and behold, white knee-socks and tan hush-puppy shoes do the trick! Of course, norms can be established in a variety of ways: top-down or bottom up, through explicit instruction or informal undirected coordination, by reference to some specialized subset of a population or to the population as a whole, or some combination of these. But however this process goes, there is no sense to be made of collective fallibility about what (response-dictated) properties make for category membership, since the collective generates the norms that determine exactly what, at any given time, these properties actually are.[6]

(3) Norm-guided property recalibration: What counts as the relevant response-dictated property for category membership will of course evolve or change as the constituting norms of the collective evolve or change. This will be obvious, I hope, from the examples given thus far. These examples should also make clear that norms for different response-dictated categories may evolve or change at different rates: the 'fashionable' is considerably less stable in this regard than the 'grammatical'. Different factors will affect the stability of such norms, including whether stability itself is considered an asset or a liability to how the particular response-dictated category functions within a practice or form of life (e.g., designing/ manufacturing/ selling clothes versus successful communication). But, again, no matter how sticky or stable the relevant norms may be, we should not let that blind us to the fact that category membership in a response-dictated category is just a function of what the norms themselves dictate.

Returning now to the topic of moral responsibility, Conventionalist Strawsonians hold that responsible agency is itself a response-dictated category. There is no independent fact of the matter about what properties of agents count as responsibility-making features; the relevant features are determined by the norms that govern a shared form of life in which this concept plays a critical role -- namely, the norms that animate our attitudes and practices of holding responsible. Conventionalist Strawsonians may have different stories about what determines these norms, or how liable they are to change or evolve over time. But the fact remains that, just as in the case of the fashionable and the grammatical, the norms themselves dictate who genuinely counts as a responsible agent.

One last point: If we return now to the epistemic question with which we began, it will be clear why Conventionalists endorse Strawson's methodological claim that attending to the attitudes and practices of holding responsible gives us insight into the nature of responsible agency. How else are we to discover the response-dictated properties that make for such agency?

1.2 *The Indicative View*

The central contention of the Indicative View is as follows:

> What it is to be a responsible agent is to have whatever it takes to be an ***objectively appropriate target*** of reactive attitudes and practices, as indicated by the underlying nature of these attitudes and practices.[7]

This view is metaphysically realist in the following sense. According to its proponents, there *is* an objective responsibility-making feature of agents that our reactive attitudes and practices are tracking (however crudely). But in order to get a bead on what this feature is, we need to focus greater analytic attention on the underlying nature of these attitudes and practices, and not just on their (possibly variable) surface characteristics. For this will give us a better sense of what it is about agents that our attitudes and practices are actually tuned into. Here, for instance, is one general thing we might notice about these attitudes and practices: they seemingly play a vital,

perhaps essential, role in certain kinds of interpersonal relationships, but not others. But why? What do these attitudes and practices contribute to such relationships?

Here is one plausible and attractive analysis that many Indicativists endorse, attributing the original thought to Strawson himself (e.g., Watson, 1987). Our attitudes and practices of holding responsible express normative demands and/or expectations. This is an attractive view on three counts: (1) it gibes with the phenomenology of reactive attitudes/practices (resentment, for instance, feels like a form of protest to someone's treating you in a way *they ought not to treat you*!); (2) it explains why these attitudes/practices are central to some kinds of human relationships, but not others – namely, those relationships that depend on our capacity to operate in the space of responding appropriately to normative demands; and finally (3) it isolates a plausible candidate for the kind of feature that makes for responsible agency, one that is vital to us precisely because it enables us to engage in this normatively enriched field of interpersonal relationships. What is that feature? Obviously, it is just the capacity for understanding and living up to normative demands/expectations as these are expressed in our reactive attitudes and practices (henceforth, I shall refer to this as a capacity for normative self-governance).[8]

But, even if we buy this analysis of what makes for responsible agency, in what sense is this really an *objective* feature of agents – one that exists independently of our reactive attitudes and practices? After all, if the analysis depends on a consideration of what these attitudes and practices require in an agent, doesn't this suggest that the truth-making property for responsible agency is itself determined by the attitudes and practices themselves? It does not – but an analogy may help make this clear.

Consider the category of things that are red.[9] It is a truism that things only look red to us because we have a certain kind of visual system. Change the visual system and you change the way things look. So it is partially thanks to the fact that we are trichromats that red things look a certain reddish way to us in standard viewing conditions; they would certainly not look this reddish way to dichromats (like dogs and elephants) or pentachromats (like pigeons). Of course, this gives us a certain advantage over dogs and pigeons: if we have normal human

colour vision, then in standard viewing conditions we can pick out the red things just by looking: they are the things that have the reddish look. Admittedly, there is a bit of a cheat here: the whole reason we have a category of <*things that are red*> is that things that have the red-making property are visually salient to us in a particular way. What red-making property is this? It is just that property – maybe a surface reflectance property, maybe something else – that makes various things look red to standard human viewers under standard lighting conditions. Still, even though it is normally or primarily detected by the way it affects our visual system (under normal viewing conditions), it is nonetheless a property (however complex) that exists in objects quite independently of our visually-guided responses to them.

Reverting to some philosophical jargon, we might say that the category relevant property of things that *makes* them red is *response-disclosed;* more precisely, it is disclosed in virtue of the sensitivities of our particular visual system. And this is important for the nature of the category so determined. As I stressed above, if we didn't have the kind of visual system we do, then likely this property wouldn't be very interesting to us; indeed, we might be completely unaware of its existence. So the very existence of red things *as* a category of things we care about or notice depends in a critical way on our (species-typical) responses to how things are in the world. To mark this fact, I will call such categories *response-disclosed categories:* categories of things whose category-relevant objective properties are (normally) disclosed by way of species-typical responses to those things. What makes these properties objective is that they exist independently of us; what makes them interesting and/or relevant to us -- indeed, worthy of systematic investigation -- is that they are capable of prompting (under appropriate conditions) a species-typical response.

As in the case of response-dictated categories, there are three phenomena worth noting in connection with response-disclosed categories (the first of these is shared in common, the latter two, marked with an asterix, are not):

(1) Individual fallibility: It is clear that particular individuals can go wrong (perhaps systematically) in their judgements as to whether something has a particular response-disclosed

property: for instance, as to whether it is red. This may be for a variety of reasons: there may be something intrinsically wrong with their red property detector (a.k.a. their visual system) – it is not functioning in the normal trichromatish way, either permanently or temporarily; there may be something non-standard about their particular viewing conditions; and, finally, there may be something odd or unusual about the object itself that prevents accurate discernment of the relevant property (for instance, in addition to its surface reflectance property, it emits a shimmery haze that significantly distorts the way it looks).

*(2) Collective fallibility: Just as particular individuals could go wrong in their judgements as to which things are red, so *everyone* could go wrong in at least some of these judgments. Again, the reasons for this may be various (though perhaps more constrained than in the individual case). So, for instance, there may be some systematically distorting viewing conditions that people can't recognize as distorting without appropriate advances in the relevant science (e.g. the effect of certain kinds of lighting conditions); or there may be something odd or unusual about particular objects that, again, is only revealed to have a distorting effect on visually discerning their properties after serious scientific investigation into the nature of those objects.

*(3) Property-guided norm recalibration: Since the property we are triangulating on by way of our species-typical response under appropriate (normal or standard) conditions is a perfectly objective feature of the world, it stands to reason that as our understanding of that property improves, so will our understanding of what constitutes the right sort of conditions for discerning that property, as well as our understanding of the kind of things that possess it. This may lead to some important recalibration in the norms we embrace for determining when our responses are accurate or adequate. Hence, for instance, we might come to recognize that our standard methods for discerning whether or not something is red (looking at it in good light) is *not* the best way of discerning whether certain objects (e.g. the ones that emit the shimmery haze) have the property that standardly makes thinks look red in good light. We are forced to develop new methods for accurately discerning the colour properties of these objects; and in

doing so, we introduce new norms for making adequate colour judgments, especially in controversial cases. Of course, this is a delicate business, reflecting the fact that the colour properties of objects are properties that we only care, and perhaps even know, about in light of our standard responses in standard conditions. So the properties we come to track in this new and improved way had better be ones that we often succeed in tracking by way of our standard responses in standard conditions. Otherwise, it may look as if we have simply changed the subject: we are no longer taking about the property of 'being red' but of something else altogether. Hence, the property-guided norm recalibration here described is likely to be a fairly conservative process, but it may nevertheless have significant consequences (especially for penumbral cases).

Returning now to the topic of moral responsibility, Indicative Strawsonians hold that responsible agency is itself a response-disclosed category. There is an independent property of agents we are triangulating on by way of our species-typical responses to agents with this property – i.e. by way of our attitudes and practices of holding responsible. But since these responses are our primary means for discerning the property in question, we could not get very far in reaching any in-depth understanding of what this property amounts to, let alone it significance for us, without a careful examination of this species-typical set of responses: in particular, we want to know when, and especially why, we take these responses to be appropriate or inappropriate. This, I trust, makes clear why Indicative Strawsonians are happy to embrace the methodological claim that is central to Strawson's paper: it provides the resources for generating a satisfying *substantive* account of what makes for responsible agency – indeed, as I noted above, for many Indicative Strawsonians this substantive account will involve, one way or another, the *capacity to understand and be governed by normative demands and/or expectations.*[10]

Of course, to be a fully satisfying account of responsible agency, more will have to be said about what it takes to have such a capacity – but, at least with this in hand, Indicative Strawsonians can point us towards a live research agenda: one that has a philosophical

dimension (perhaps we need to say more about what it means to have such a capacity at a conceptual level – more on this below) and an empirical dimension (we certainly need to understand more about the (undoubtedly complex) psychological features of agents that are required to support such a capacity). In addition, Indicative Strawsonians may promulgate the importance of this research agenda on normative grounds – for if we have a better understanding of what it *takes* to have such a capacity, then we may have grounds for recalibrating the norms we embrace for judging whether, and to what extent, various agents are responsible (e.g. we may abandon the intuitively reasonable 'quality of will' test for responsible agency, at least for psychopaths, if it turns out that their quality of will is simply irrelevant for assessing their capacity for normative self-governance).

Section 2: Relative merits of the standard views

As I have emphasized in Section 1, the standard views are committed to defending Strawson's central methodological insight: that philosophers only make progress on understanding what it takes to be a responsible agent by focussing on our attitudes and practices of holding responsible. But they do so in starkly different ways. The Conventionalist View is metaphysically stipulative in so far as it insists that 'being responsible' is a response-dictated property of agents; specifically, what it takes to be a responsible agent is determined by whatever norms are in place that govern a community's reactive attitudes and practices. By contrast, the Indicative View is metaphysically realist in so far as it insists that 'being responsible' is a norm-independent, response-disclosed property of agents; specifically, what it takes to be a responsible agent is simply made salient to us by considering the general shape of our attitudes and practices of holding responsible. For the properties presupposed by, and generally discerned in virtue of, our proneness to such attitudes and practices are the very properties pre-requisite for meaningfully engaging in the kind of norm-governed interpersonal interactions that we pre-theoretically view as properly reserved for responsible agents.

I turn now to assessing the relative merits of these two views. But in order to do so I need to add one other dimension of Strawson's agenda in refocussing our attention on the attitudes and practices of holding responsible. It involves his preoccupation with a substantial credit/discredit notion of merit or desert.[11]

Recall that Strawson's aim in "Freedom and Resentment" was to intervene in a dialectic that he thought was going nowhere. The disputing parties – his so-called 'optimists' and 'pessimists' – seemed each to be stuck in an implausible rut, driven there by their unwillingness to give up on one or another entirely reasonable desideratum that an account of responsible agency should meet. The optimists, on their side, were bound and determined to defend an account of responsibility that could justify our practices of praise and blame (as deterrent measures) without assuming that agents must exercise some spooky 'agent-causal' power in order to be effectively regulable, and thus responsible, for what they do (i.e. an account of responsibility that is compatible with the metaphysical thesis of determinism). The pessimists, on their side, were equally vociferous in explaining why such an account would not do: the optimists' approach might succeed in justifying our practices of praise and blame if mere 'behaviour regulation' were the point and purpose of such practices. But praise and blame are normatively far more significant than this: they express a moral assessment of their recipients as the kind of creatures who, in a normatively substantial sense, *deserve* or *merit* our praise or blame. In the pessimists' eyes, such normatively substantial desert could hinge on nothing less than agents being 'ultimately' responsible for their activities, leading them to insist on an account that makes essential reference to agents' having and exercising the very agent-causal power the optimists disdain (i.e. an account of responsibility that is manifestly *in*compatible with the metaphysical thesis of determinism).

The problem, as Strawson sees it, is that a fully satisfying account of responsibility must meet both desiderata here at issue: (1) the optimists' desideratum, that such an account should avoid invoking spooky agent-causal powers as a feature of responsible agency (even better, according to Strawson, it should avoid the baneful topic of metaphysical determinism

altogether); and (2) the pessimists' desideratum, that such an account should satisfy our very deep intuition that praise and blame are only justified in so far as there is a normatively substantial sense in which so-called 'responsible' agents merit or deserve these sorts of responses; hence, a normatively substantial sense in which their activities redound to their credit or discredit. Strawson's aim in refocussing our attention on the reactive attitudes and practices is thus importantly intended, at least in part, to satisfy the pessimists' desideratum. For, as his exhaustive discussion of these attitudes and practices makes clear, he agrees with the pessimists that a normatively substantial notion of desert is part and parcel of what animates them. Hence, any proposal for the responsibility-making property of agents that we arrive at by adopting Strawson's methodological advice, it had better be the kind of property that can bear this kind of normative weight. Otherwise, we are back once more in the same dispute, facing the same unpalatable alternatives that Strawson aimed to leave behind.

So in this section, I consider the relative merits of the Conventionalist View and the Indicative View in regard to how well they fare on the pessimists' 'desert test'. We have seen how each of these views specifies the kind of property that makes for responsible agency. The question we face now is whether either view succeeds in convincing us that having such a property is really all that it takes to be *genuinely deserving* of our reactive responses in (something like) the normatively substantial credit/discredit sense that Strawson's pessimists had in mind?[12]

I begin with the Conventionalist View, but fear I give it short shrift. For this view can only succeed at passing the pessimists' desert test by discrediting the independent standing of our desert intuitions in providing some check on theorizing about responsible agency. Obviously, Strawson's pessimists would not be happy about this – and I am not sure Strawson would be either (this, of course, is open to debate). Still, some may see this a winning strategy -- indeed, a strategy Conventionalist interpreters of Strawson attribute to Strawson himself; so it's worth explaining how it goes.

Begin with the thought that our desert intuitions *do* provide some kind of independent check on theories of responsible agency. From this perspective, the Conventionalist View clearly fares poorly on the pessimists' desert test. After all, Conventionalists hold that responsible agents are just those agents that are deemed to be 'appropriate' targets of our reactive attitudes and practices, given the currently accepted norms that govern those attitudes and practices. We have already noted that potential norm variability is a feature of this view; and, indeed, there is strong evidence that such variability exists across cultures and historical periods (for discussion, see Lacey, 2016). For instance, the ebb and flow in the popularity of animal trials in different times and places seems to point to shifting norms surrounding who (or what) we think counts as an appropriate target of reactive attitudes and practices.[13] But surely, the desert critic will say, we *ought* to disavow this practice as completely wrongheaded. Animals are not the kind of beings that *deserve*, in any substantive normative sense, to be held responsible for what they do – hence, to be tried, judged and sentenced in a court of law: the very idea is a farce. Hence, the Conventionalist View fails to provide a fully satisfying account of what makes for responsible agency. For even if some set of currently accepted norms actually succeeds in targeting agents that happily pass the intuitive desert test, this is by no means guaranteed by the view; and that is enough to reject it.[14]

But proponents of the Conventionalist View have an obvious response to this charge. Anyone who gives such credence to our ordinary intuitions of genuine desert is clearly presupposing a substantial independent account of what it 'really' takes to be a responsible agent, quite apart from the norms that govern our reactive attitudes and practices. But this is just what the Conventionalist denies. So the idea that there is some self-standing intuition of desert that can do this critical work is simply illusory. Of course, Conventionalists will agree that it can seem *to us* as if we have a handle on what it is to be genuinely deserving of reactive attitudes and practices in some norm independent sense; but, in their view, this would simply show how deeply we have internalized our own proprietary cultural norms. In the end, thorough-going Conventionalists will simply deny that there is any non-question-begging way

of subjecting their view to the pessimists' desert test; for it clearly presupposes the very thing they reject.

What is there left for the desert critic to say? Only this: that these two things tend to stand and fall together. If you accept the idea that it's reasonable to subject an account of responsible agency to the pessimists' desert test, then, barring very strong arguments to the contrary, you are already inclined to reject a metaphysically stipulative approach to this topic.[15] You're inclined to suppose there must be some legitimate external check on our attitudes and practices of holding responsible – namely, a norm-independent feature of agents that makes them responsible. In effect, you're anxious to part company with proponents of the Conventionalist View and endorse an intuitively more plausible, albeit theoretically respectable, alternative. The Indicative View might just fit the bill. Indeed, prominent defenders of this view insist that this is its main attraction and rationale. Jay Wallace, for instance, argues that it is a primary concern with 'fairness' that leads him to spell out, in substantive objective terms, the property of agents that makes them *intuitively deserving* of our reactive responses (Wallace, 1994). So it seems right and proper to explore how views of this Indicative type fare on the desert test.

I said earlier that a variety of accounts fall under this general Indicativist umbrella, with different accounts spelling out the nature of the responsibility-making property in a range of different ways.[16] But since the most persuasive of these begin from a functional analysis of our reactive attitudes and practices in terms of expressing normative demands/ expectations, I focus here on views that highlight a responsible agent's *capacity* to understand and live up to such demands/ expectations – what I earlier called a capacity for normative self-governance. Such views intuitively have what it takes to pass the desert test. As Wallace himself argues:

> "… once we correctly understand the stance of holding people responsible in terms of the reactive emotions, we will be able to see that *the condition that makes it fair* to adopt this stance is not freedom of the will in the strong sense; rather it is the kind of normative competence in virtue of which one is able to grasp moral reasons and to control one's behaviour by their light" (Wallace, 1994, p.: 16, my emphasis).

So taking my cue from Wallace – and, indeed, a number of like-minded Indicativists – I will treat the capacity for normative self-governance as nothing more than a kind of ***reason-responsive capacity***, specifically the capacity to track and respond to moral reasons. On this view, agents can only deserve praise and blame (or other reactive responses) in the substantial normative sense in so far as they possess the (general) capacity to track and respond to (moral) reasons in the circumstances in which they act; for only then does it seem fair or fitting to 'hold them responsible'.

Of course, proponents of this view now owe us a *substantive* account of what it means to possess the requisite capacity. And by this, I do not mean they owe us an empirical account of what it would take to realize such a capacity in psychological terms. Rather, they owe us a metaphysical account of what, in their view, a capacity *is* such that their capacitarian approach to responsible agency really does give them what they need to pass the desert test.

So what does it mean to possess a capacity, on the Indicativist view? This, too, is a matter of some controversy. But perhaps the most explicit account is simply this: to possess a capacity is to possess a *disposition*, or set of dispositions.[17] Further, according to the "new dispositionalists" (the term comes from, Clarke, 2009), the most plausible metaphysical analysis of what it means for something to possess a disposition (or set of dispositions) is for a certain counterfactual claim to be true of that thing (Fara, 2008; Manley & Wasserman, 2007; Smith, M., 2003; Vihvelin, 2004). To wit:

- it *would* manifest certain characteristic behaviours *under a range* of characteristic conditions.

Further, it is presumed that possessing a disposition in this sense is to be explained in terms of the thing's possessing some underlying categorical (or intrinsic) property -- a *structural feature* of the thing -- that makes the counterfactual claim true; that explains why it manifests certain characteristic behaviours under a range of characteristic conditions. 'Fragility' is a good model for what's intended here. To wit:

- A glass is fragile iff it is *inherently so structured* that under a range of characteristic conditions (droppings, strikings, throwings, etc.), it will shatter.

Likewise,

- A person is reason-responsive iff she is *inherently so structured* that, under a range of (fairly open-ended) conditions she tracks and responds to reasons.[18]

There are two points in connection with the dispositional analysis of capacities that bear emphasis. First, just as it is not luck or happenstance that a fragile glass breaks when it drops, it is not luck or happenstance that a reason-responsive person tracks and responds to reasons. Both are (naturalistically) explicable in terms of some structural feature inherent in them. But, secondly, merely possessing the requisite intrinsic property -- hence, the disposition/ capacity -- does not *entail* that the thing or person *will* manifest the characteristic behaviour under characteristic circumstances. Fragile glasses don't always break when dropped; reason-responsive agents don't always respond to the reasons.

This second point is essential for understanding how responsible agents could *ever* be such as to deserve some negative reactive response in light of what they do – i.e., blame, resentment, indignation, censure, etc. For, on the view we are considering, it has to be such that the agent possesses the requisite reason-responsive capacity but (culpably) fails to exercise it. This means she must possess the requisite capacity to track and respond to the relevant moral reasons *in the circumstances in which she acts* – i.e. there is nothing blocking her *in those very circumstances* from exercising her capacity.[19] She simply fails to exercise it. That is to say, there is no excuse – no mitigating or perturbing factor – that might *explain* her failure in a way

that lets her off the hook (e.g., she was coerced, she was understandably distracted, she was 'not herself' due to uncontrollable anxiety or whatever). No, she simply fails to exercise the capacity she has, just as a glass might simply fail to shatter when dropped (there is no pillow to prevent it from breaking, no lucky catch, no sudden change in the earth's gravitational field, etc.).

We are now in a position to see why passing the pessimists' desert test is actually deeply problematic for the reason-responsive capacitarian view of responsibility. For consider: In the case of the glass, we simply accept the fact that some chance causal factor explains why it didn't break in circumstances under which it normally would (given its fragility). But in the case of the person, this can't be what we think. For if we genuinely believe that some chance causal factor explains why she didn't respond to the reasons in circumstances under which she normally would (given her capacity for reason-responsiveness), we are surely committed to thinking she does not *deserve* blame or censure for what she did. After all, failure to exercise her capacity was not *down to her* in any normatively significant way.

Aha, the Indicativist may reply. This precisely highlights the significant difference between the two cases. For, with regard to people, we often do believe there are factors at work that both explain someone's failure to exercise her capacity and that are down to her in a normatively relevant way. For instance, maybe she was simply too lazy to bother thinking about the relevant reasons. The problem with this very natural thought is that it just seems to move the deckchairs around a bit. For we can simply ask: did she have the capacity to overcome her inherent laziness in the circumstances in which she acted? If she didn't, then it seems we are blaming her for something she couldn't do anything about. And if she did, then why did she fail to exercise that capacity? Of course, there must be some explanation for her failure in this case – presumably, some chance causal factor intervened. But if that's the case, then, again, we seem to be blaming her for something that was not down to her in any normatively significant way.

Is there really no way out of this bind – no way to explain how our reason-responsive agent could fail to exercise her capacity in a way that is *genuinely* down to her, thereby making

her *genuinely* deserving of blame? Well, of course, there is one possibility I have not yet canvassed. She deserves our blame because she had some *special causal power* to exercise her capacity that she simply chose not to use. I trust it is clear why we don't want to embrace this particular option. But what actually is left, given the dispositional analysis of what it means to possess a reason-responsive capacity? Reason-responsive Indicative theorists may twist and turn to find some way out of this "hard problem of responsibility", but I think the prospects are dim.

So where have we got to so far? In assessing the relative merits of Conventionalist and Indicative Views of responsibility, it seemed that only an Indicative View showed some initial promise of passing the pessimists' desert test. But the most attractive family of Indicative Views – those that associate responsible agency with a capacity for recognizing and responding to (moral) reasons – doesn't seem to deliver the goods in the end.[20] So either we give up on meeting the pessimist's desideratum – i.e. give up on the idea that so-called responsible agents *ever* deserve our reactive responses in a normatively substantial sense; or we find a new way forward. My aim in the rest of the paper is try for the new way forward, beginning with an inquiry into the apparent root of the Indicativists' failure: the dispositional conception of what it means to possess a capacity. [21]

Section 3: The metaphysics of 'intelligent capacities' – a.k.a. skills

The dispositional model of capacities has a distinguished reputation, at least amongst naturalistically inclined philosophers. For, as we have seen, dispositions seem to be straightforwardly explicable in terms of some underlying categorical (even if complex) physical property possessed by the thing in question. Still, despite the attractions of this model, there are certain commonsense observations we are inclined to make about an important subset of our capacities that seemingly mark them off from 'mere' dispositions. These capacities are many and varied, but Gilbert Ryle dubbed them all "intelligent" in order to highlight a key aspect of their distinctive nature (1949, pp. 42-43). More colloquially (and also following Ryle), we may simply characterize these as 'skills'. Rylean examples include: *playing chess, mountaineering,*

driving a car, *target-shooting*, *constructing arguments*, *calculating sums* – and (here adding to Ryle's list) *responding to reasons*. In this section, I build on Ryle's insights to develop an alternative model of the underlying metaphysical structure of such capacities, beginning first with the commonsense observations that lead us away from a more straightforward dispositional view.

First commonsense observation: intelligent capacities come in degrees. How might the standard dispositional model of a capacity accommodate this fact? As we have seen, the model holds that having a capacity is explicable in terms of how an agent is structured at a given period in time; she is so structured that she would manifest certain characteristic behaviours under certain characteristic conditions. So to change how she is structured at a given time *is* to change or alter what capacities she has. Hence, the straightforward observation that a particular capacity may come in degrees is in *prima facie* tension with what the model allows.

The dispositionalist can surely handle this kind of objection. After all, it looks like a merely verbal dispute, turning on how we individuate capacities: the dispositionalist does it in a very fine-grained way; commonsense is more rough and ready. In any case, this observation marks no significant difference between intelligent capacities and mere dispositions. For instance, it seems entirely reasonable to say that one glass can be more or less fragile than another because of some difference in their respective underlying structures; or, indeed, that the *same* glass can become more or less fragile, precisely because of some change or alteration in its underlying structure. So too for agents: for instance, they can be more or less reason-responsive thanks to some difference in their underlying (presumably psychological) structures; or, indeed, the same person can become more or less reason-responsive thanks to some change or alteration in her underlying psychological structure.

Hence, the dispositional model accommodates a second commonsense observation: that individuals change with respect to how skilful they are --i.e. to what degree they possess a given (commonsense) capacity. There is nothing mysterious here. The dispositional model simply commits us to the following view: if we say a person 'has become more reason-responsive, *all*

we could possible mean by this remark is that they are *now* so structured that they *now* would respond to the reasons in a greater range of circumstances than they did previously.

This all seems fine as far as it goes. The time-slice dispositional model of what it means to have a capacity can accommodate the commonsense observations that capacities come in degrees, and that a thing can change over time with respect to what degree of capacity it currently possesses. And yet these observations are linked, I think, to a deeper phenomenon that simply gets obscured by the atemporality of this model -- namely, the sense in which we think of intelligent capacities as ***essentially developmental*** in nature.[22] Ryle nicely captures this point in his discussion of the difference between skills and habits:

> "It is … tempting to argue that competences and skills are just habits. They are certainly second natures or acquired dispositions, but it does not follow from this that they are mere habits… It is of the essence of habitual practices that one performance is a replica of its predecessors. It is of the essence of intelligent practices that one performance is modified by its predecessors. The agent is still learning" (Ryle, 1949, p.: 30).

Ryle's observation provides the key insight on which an alternative 'skill-based model' of intelligent capacities can be elaborated. For, unlike mere dispositions, it's in the very nature of these capacities, not just to change, but to develop over time; and they develop over time precisely because of the way an *intelligent agent* interacts with her environment -- because she probes, tests, explores different ways of doing things, and adjusts what she is doing in light of the feedback she receives. As Ryle says, "the agent is still learning".

Building on this observation, we can now identify three further features of intelligent capacities that distinguish them from mere dispositions:

(1) Intelligent capacities ***take work, or 'practice', to develop***, where the kind of work in question involves feedback from the environment. As already noted, Ryle's examples include such things as mountaineering and target-shooting; but we can expand his list in obvious ways.

Think, for, instance of our linguistic skills -- or other essentially social skills, like becoming an adept folk-psychologist (a.k.a 'reader of other minds').[23]

(2) Intelligent capacities also ***take work, or practice, to sustain***. Sadly, as we all know, skills can get rusty through disuse; they can decay and disappear (for instance, consider that foreign-language you once spoke so well as a child, now frustratingly elusive both in comprehension and production).

This feature of intelligent capacities is, I think, particularly obscured by a philosophical tendency to embrace the dispositional model of capacities. For mere dispositions, at least of the kind we have been considering (e.g. fragility), are relatively stable features of things; indeed, it generally takes work to *change* these dispositional features -- e.g., heating glass to temper or strengthen it. By contrast, it takes work to *sustain* our skills, work that may be more or less invisible to us, but work nonetheless -- e.g. we practice the language we currently speak *all the time*, day in and day out, alone and in the company of others (imagine how fit we would be if only we devoted this much time and effort to exercise!).

Of course, skills can be more or less resilient to decay, though this is hard to measure given the confounding variable of constant practice. Nevertheless, it does seem true that, once having acquired a skill, getting it back 'in shape' is not as difficult as acquiring it in the first place. Still, despite all these provisos, it seems clear that skills are ***essentially fragile*** in a way that mere dispositions are not.

(3) The kind of environmental feedback we need to develop and sustain our intelligent capacities will partly depend on the nature of those capacities.

Some capacities do not require *specifically social* feedback in order to develop and sustain them. Think here of target-shooting, or mountaineering. Of course, this does not mean that social feedback (e.g. in the form of teaching) isn't extremely helpful in developing and sustaining such skills -- indeed, it may even be practically indispensable given the kind of creatures we are.[24]

But some capacities (e.g. our linguistic capacities) are essentially dependent on social feedback in order to develop and sustain them, precisely because they are social skills: they are skills that involve some competence at navigating within an interpersonal norm-governed environment. The skills we need to develop are therefore skills in understanding and complying with a set of mutually shared and interpreted norms -- often a matter of on-going negotiation and adjustment, particularly as the norms themselves are liable to change in response to changing social and environmental circumstances. Our capacities to engage in shared norm-governed activities thus depends on our being tuned into how others respond to us, and we to them, thereby all doing our part to make and sustain the very norms that make such activities possible.

Again, these differences between *kinds* of intelligent capacities, not to mention the differences between such capacities and mere dispositions, simply go unnoticed on the standard dispositional model.

In sum, I think it is clear that the skill-based model of intelligent capacities is essentially unlike the disposition-based model with which we began. It calls for a fundamentally different kind of metaphysical analysis of such capacities: one that is explicitly dynamic, inter-temporal and inter-personal rather than static, atemporal and intra-personal. To mark this fact, I will call my view a metaphysically 'constructivist' account of capacities, rather than a straightforwardly realist one. This label, though, is a bit misleading. For I don't deny that there are substantial objective features of agents that underwrite such capacities. However, given the accordion-like nature of these capacities, the features in question must be such as to explain the agent's characteristic *developmental sensitivity* to the environmental feedback she encounters.

Section 4: The Scaffolding View -- A metaphysically 'constructivist' account of responsible agency

Having sketched an alternative picture of (intelligent) capacities, I now return to the question of responsible agency. In this section, I explore and defend a non-standard Strawsonian view of what it takes to be a responsible agent. I call it the 'Scaffolding View'. This view is certainly not metaphysically stipulative, in the manner of the Constitutive View. But nor is it metaphysically realist, at least in the manner of the Indicative View. So I call it 'metaphysically constructivist' in precisely the sense I used above in connection with skill-based account of intelligent capacities.

The Scaffolding View is not some middle compromise between these two standard Strawsonian views. On the contrary: It leans much more heavily in a realist direction, building on the central insight of what I claimed was the most persuasive version of the Indicative View. This insight, as discussed in Section 2, is that responsible agency depends on having a capacity for recognizing and responding to moral reasons. And though I think reason-responsive Indicativists make good arguments in defence of this claim, an advantage of the Scaffolding View is that it makes even clearer why this claim should seem compelling.

So, the key difference between the Scaffolding View and the (reasons-responsive) Indicative View is that it relies on the skill-based model of what it means to have a capacity for recognizing and responding to reasons.[25] To briefly recap, the skill-based model presents such intelligent capacities as distinctive in the following ways: (1) they are an accordion-like feature of agents that essentially come in degrees; (2) they are susceptible to development and decay -- in that sense, they are fragile; (3) developing and sustaining such capacities depends in part on certain internal (objective) properties of the agent; but (4) it also depends, in very large part, on getting the right sort of feedback from the environment, where such feedback may be essentially social in nature.

In this section, I focus on the implications of these last two aspects of intelligent capacities for the capacity here of interest: the capacity to recognize and respond to moral reasons. That is to say, my focus will be on the conditions, both internal and external, that are necessary for developing and sustaining such a 'fragile' capacity. This is not, by any means, a

fully developed account. My aim is, rather, to sketch the basic ideas at work in the Scaffolding View, and to try and show what makes these ideas both natural and appealing.

I begin in reverse order, with a consideration of the type of feedback essential to developing and sustaining the capacity to recognize and respond to moral reasons. Although I think this is true of any reason-responsive capacity, I claim, in particular, that the capacity to recognize and respond to *moral* reasons is an essentially social skill, requiring social feedback to develop and maintain. For moral reasons are themselves concerned with our interpersonal relationships -- specifically, with how we ought to treat and regard one another in the context of those relationships (though of course moral reasons may extend beyond this interpersonal domain as well). But how we are to treat and regard one another is something we work out together, developing in community with others a shared set of norms that are in their nature never static, but continuously subject to re-negotiation in light of changing epistemic and material circumstances. To recognize and respond to moral reasons is thus to be sensitive to such norms: to understand what they demand of us and to govern our actions accordingly -- or, of course, to challenge such norms if we think they are normatively objectionable. But this is an on-going project requiring continual social feedback, both because the norms themselves are subject to socially negotiated challenge and change, but also because the norms themselves are invariably complex and open to socially negotiated interpretation in the demands they make of us.[26]

If we accept this general picture of why we rely on feedback from one another to develop and maintain our capacity to respond to moral reasons, especially in an interpersonal context, we are now in a position to ask what form this feedback takes. In particular, what role do our *attitudes and practices of holding responsible* play in this process? Are they part and parcel of the necessary feedback we receive -- or is their role in our interpersonal relationships something quite distinct?

A key move of the Scaffolding View of responsible agency is made here. It is to insist that our attitudes and practices of holding responsible play a critical role in developing and

sustaining our capacity to recognize and respond to moral reasons. They are in that sense rightly viewed as "proleptic" attitudes and practices -- attitudes and practices that help call forth the very thing the attitudes and practices presuppose.[27] This sort of view can sound paradoxical, or even illegitimate, without a skill-based account of our capacities in the background. But once that account is in place, I suggest that the proleptic view is an obvious one to hold.

Why is that? Again, I build on a central insight of Indicative theorists. It is that reactive attitudes and practices are a form of moral address, expressing normative demands and expectations. But Indicative theorists do not make enough of this point, in my view. For addressing someone is not generally a unidirectional activity. We don't simply talk at people when we address them; we expect them to respond to us in some way appropriate to how we have addressed them. So, to zero in on the negative attitudes for the moment (e.g. the blaming attitudes of resentment or indignation), if these are a form of moral address, then they are not in their nature merely backward-looking 'reactive' attitudes; they are also forward-looking 'evocative' attitudes (as we might say),[28] *calling for* a particular response from the putative wrongdoer: for instance, to explain and/or justify what they have done; and where that fails, to acknowledge how they have failed to track and respond to the reasons that ought to govern their behaviour, and to commit to doing better in the future.

Thus, on the proleptic view, we can think of our so-called 'reactive' attitudes as properly embedded in trajectories of reactive exchange (McGeer, 2012, 2014) this is what gives them their point and purpose. Your blame, for instance, calls for an appropriate 'reactive' response from me -- e.g. recognition of my moral blunder. Presuming that I have no excuse for my bad behaviour, such recognition involves granting that there were moral reasons for behaving otherwise to which I was not sensitive. But your blame demands that I go beyond mere recognition of this fact; it demands that I regret it, that I feel sorry about it, and that I am consequently motivated to take responsibility for doing better in the future, thereby *expanding my capacity* for tracking and responding to the reasons that there are. And I show all this by way of *my* reactive responses to your reactive blame -- i.e., by way of my guilt, shame, remorse,

contrition. And then you, all going well, will have a natural reactive response to these reactive attitudes of mine as a consequence of what you properly take them to express: viz., that I have understood what motivated your original complaint, that I take responsibility for it, and that you can reasonably expect me to better in the future. Thus, in supportive and hopeful acknowledgement of my increased capacity for tracking and responding to moral reasons, your angry resentment is appropriately replaced by the reactive attitude of forgiveness; whereupon I, reinforced by your reactive forgiveness, become more confident in my own enlarged capacity to track and respond to moral reasons. And so it goes.[29]

This, of course, characterizes an 'ideal' type of reactively trajectory -- ideal, especially from the blamer's point of view since it vindicates the blamer's original complaint. But there are other 'ideal' types of reactive trajectory, given a proleptic understanding of such reactive exchanges. For instance, imagine that you blame me, as before. But now instead of reacting with guilt/ shame/ remorse/ contrition, I am angrily indignant with you in turn. What does this sort of reactive response express? At the very least, that I don't agree with your normative assessment of my behaviour. I don't see that I did anything wrong -- and, indeed, I think *you* are wrong in judging otherwise, and even more wrong to call on *me* to reject my earlier behaviour and 'commit to doing better'. Even more, my indignation calls on *you* to change your attitudes and your (blaming) behaviour. In short, we are now engaged in a substantial disagreement over the moral reasons that ought to govern our behaviour; and this implies that one of us (at least) needs to enhance their capacity to track and respond to the reasons that there are. And so we 'enter into a negotiation' about who needs to pull up their socks, likely with some passion on either side (reflecting our concern, I would suggest, with what we take to be at stake). An ideal scenario involves coming to some agreement over this, where agreement involves some shared acknowledgement of the reasons that there are, some shared understanding as to who has failed to track those reasons and why (of course, this may be both of us), and finally some shared understanding of what we owe to one another, namely a

commitment to do better going forward. In this way, at least one, but likely both of us have enlarged our capacity for tracking and responding to the reasons that there are.

In short, on the proleptic view, our reactive attitudes and practices are mistakenly characterized as simply backward-looking responses to one another's attitudes and behaviour. Their significance lies instead in their powerful forward-looking character. For, as we have seen, they play an active role in shaping and regulating our future attitudes and behaviour by way of developing and sustaining our capacity for (socially negotiated) moral-reasons-responsiveness. Hence, they play a central and critical role in scaffolding what it takes to be a morally responsible agent.

Finally, I turn to the question to which all of this has been leading -- although (thankfully) I will here be brief. The question is, what objective underlying feature must agents possess in order to be genuinely responsible; hence, to be appropriate targets of our reactive attitudes and practices? And this, as we have seen, is equivalent to the question: what objective underlying feature must agents possess in order to have a capacity for moral-reasons-responsiveness in the skill-based sense?

On the view I am sketching, the answer is simply this: what they need to possess -- indeed, all they need to possess -- is a susceptibility to the scaffolding power of reactive attitudes, experienced as a form of moral address. Notice, first, that this is a substantially more modest requirement on responsible agency than was suggested by the Indicative View. For, on that view, at the time of their purported misdeed, agents must possess an atemporal, purely intra-personal, disposition-based capacity for responding to the reasons. This, as we have seen, is a disposition to be in-the-moment responsive to the reasons present at the time of their action -- a disposition that makes their failure to respond to those reasons both puzzling and apparently blameless. By contrast, the Scaffolding view holds that agents must possess an inter-temporal and (essentially) interpersonal skill-based capacity for responding to the reasons -- i.e. a capacity that involves having whatever it takes to be *sensitizable* to the kind of reasons present

at the time of their action, in part by way of the exhortatory effects of (ex-post) reactive scaffolding.

Still, even though the Scaffolding View defends a more modest picture of what it takes to be a responsible agent, it generates substantive desiderata for any fully worked out account of the underlying features that enable such agency -- presumably one that is rich in psychological detail. For such an account will have to focus on those features of agents that make them peculiarly sensitive to the scaffolding power of reactive attitudes and attitudes and practices. We want to know, what is it about agents that enables them to be *aware* of the normative features of these attitudes and practices -- i.e. that they express normative demands/expectations, and that that they comment on the normative acceptability of people's attitudes and behaviour. But we also want to know what it is about agents that makes them *peculiarly responsive* to the normative demands/ expectations therein expressed? And here I don't mind speculating, bye the bye, that typical human emotional sensitivities will play a very important role in the detailed psychological account we eventually construct. But I leave this issue for another day.[30, 31]

Section 5: Returning to the problem of desert

I conclude this paper by addressing a last outstanding concern. I said in section 2 that an account of responsible agency will only be fully satisfying if it passes the pessimists' desert test. That is to say, the account of responsible agency we end up with must be such as to satisfy our intuitive sense that the agents (including ourselves) that we regard as appropriately targeted by our reactive attitudes and practices are agents that *genuinely deserve* those reactive responses: genuinely deserve our gratitude, resentment, indignation, shame, remorse and forgiveness. How does the Scaffolding View fare on this score?

For historical reasons, the problem of desert has mainly been addressed in connection with blame (praise is sometimes mentioned in this context as well, but mostly as a sideshow). As I suggested above, this narrow preoccupation with blame has tended to distort our

understanding of the general shape of our reactive exchanges more broadly considered, obscuring their forward-looking 'call and response' structure (Macnamara, 2013). In particular, it has perpetuated the pessimists' entirely backward-looking focus in the way they have framed the desert question. It is framed in such a way that it seems to be solely concerned with a specific act the agent performed in the past (we might call this the 'reactive instigating' act) and, in particular, with how that specific act was produced -- i.e., it is framed as a question that concerns the metaphysical status of the agent's action in the context of a field of options that were, in some sense, open to her *in the very circumstances* in which she acted. This act-focussed metaphysical framing is reinforced by an entirely natural thought that reasonably accompanies any legitimate reactive response: 'you could have done otherwise!'.

The Indicative View, as we have seen, is traditionalist in so far as it does not break with this act-focussed metaphysical framing of the desert question. What it does -- *all* that it does -- is attempt to answer that question by replacing the pessimists' demand for some spooky causal power with the snazzy compatibilist reassurance that the agent is *actually* so structured (psychologically and, ultimately we presume, physically) such that she *would have acted otherwise* in a robust range of circumstances very like the ones she was actually in. Indeed, on this compatibilist story, it seems the *only* thing that could have prevented her from doing otherwise 'in the actual sequence', given this structure, was some unfortunate glitchy causal factor out of her control. Notice, then, that both types of account, compatibilist and incompatibilist, succeed in their primary objective of preserving the truth of the claim, "you could have done otherwise", where that claim is taken to register *some* kind of abstract metaphysical possibility. The difference between them lies in the kind of abstract metaphysical possibility they defend: in the pessimists' case, it is a robust *causal-originating* kind of possibility; whereas in the Indicativists' case, it is a more anaemic (and so naturalistically-friendly) *causal-modal* kind of possibility -- the agent is suitably located in a space of possible worlds. But ludicrous as Strawson (or we) may find the pessimists' venture into the realm of spooky agent-causal powers, at least their strategy has one thing going for it. It makes *blaming*

the agent for what she did seem intuitively reasonable. The same cannot be said for the Indicativists' strategy: the agent they so characterize, who (glitchily) acts in the way she does, seems more deserving of our consolation than our condemnation. And if Indicativists continue to insist that they have provided the *only* plausible metaphysical story under which our blame could be regarded as 'fitting', then so much the worse for our metaphysical stories.[32] In the end, as Strawson says, they seem to provide nothing more than "a pitiful intellectualist trinket for a philosopher to wear as a charm against the recognition of his own humanity" (Strawson, 1962, p.: 24).

So what human condition can take the place of these arid metaphysical possibilities beloved of intellectualist philosophers? How can we reframe the desert question in a way that does not reflect this act-focused metaphysical obsession? (I hope that my phrasing does not sound excessively polemical!) These questions return us to the thought I said we might reasonably have in connection with the doings of responsible agents -- viz., "you could have done otherwise". Our problem, then, is to understand this thought, not as a claim about abstract metaphysical possibilities, but as a thought representing some actually realizable condition nonetheless. What condition could this be?

The Scaffolding View has a ready answer to this question. The condition is one of possessing the kind of skill-based capacity that comes in degrees and that takes work to develop and sustain. "You could have done otherwise" thereby purports to capture a truth -- the truth of possessing such a capacity. The thought (and even more likely, the heatedly voiced claim) is, of course, *occasioned by* what an agent actually does. But it takes a philosophically biased ear to hear the practical concern thereby expressed as one that is exclusively focused on the actual (in-the-moment) doing of an act, let alone on its metaphysical underpinnings. The practical concern is rather with situating this particular act in the larger framework of an agent's potential doings that we think she can be brought to achieve, if only by our encouragement and insistence.[33]

This suggests the following view: In thinking and (more likely) saying, 'you could have done otherwise", we are rightly adopting a stance towards agents that honours their

developmental potential. We are effectively telling them, "you have what it takes!" -- 'otherwise (suppressed premise) why would be bother taking you to task in this way?!' Of course, it needs to be added that *in so taking agents to task*, "you could have done otherwise!" is generally understood to have a perlocutionary force that goes beyond the mere reporting of some condition we take the relevant agents to be in (consider: "you could have done otherwise!" is rarely said in the calm cool tones of a philosopher reporting some abstract metaphysical possibility). Rather, we are *exhorting* agents to do better. And in so exhorting them, we thereby do our bit to sensitize agents to the reasons they earlier failed to track, and thereby do our bit to help realize the developmental potential that our claim, "you could have done otherwise" attributes to them. This makes such exhortative claims rather special from a speech-act perspective -- they aim to realize the condition they report; but they are hardly metaphysically mysterious for all of that.[34]

So what to say about 'genuine desert'? It should be clear by now that the Scaffolding view insists on reframing the desert question. It argues that we should not be committed to framing this as an act-focused metaphysical question in the way that philosophers have so often done. Rather, it should be framed as a person-focused capability question.[35] With that framing in mind, the desert question becomes something like the following: do agents, possessed of a skill-based reason-responsive capacity, deserve to be subject to attitudes and practices that in their nature exhort them to do better? Do they deserve to get the feedback they need from us (and we, by the way, need from them) to develop and sustain such a capacity? Or are our efforts at exhortation inappropriately targeted on them? There may not be universal agreement on an answer to these questions: perhaps some may feel that encouragement and exhortation is for busybodies. But at least it widens the options of what we might mean by 'genuine desert', while yet embracing the central thought that many retributivists associate with the moral imperative to give wrongdoers what they deserve -- namely, it respects who they are as persons. The thought here added is simply that we fail to respect others as persons, so far as we fail to give them the feedback they need to develop and sustain the reasons-responsive capacities that

underwrite their status as persons. Admittedly, the notion of desert here defended is not the austere metaphysical notion beloved and celebrated by the pessimists. But it seems to me a normatively substantial notion of desert nonetheless, and one that is well worth celebrating in its own right.[36]

Victoria McGeer
Princeton University, Princeton NJ
USA
Australian National University, Canberra ACT
Australia
vmcgeer@princeton.edu

NOTES

(Inserted below under references)

REFERENCES

Ayer, A. J. (1980). Free will and rationality. In Z. v. Straaten (Ed.), *Philosophical Subjects*: Oxford University Press.

Byrne, A., & Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences, 26*(1), 3-21. doi:10.1017/S0140525X03000013

Chisholm, R. M. (1982). Human Freedom and the Self. In G. Watson (Ed.), *Free Will* (pp. 24-35). Oxford: Oxford University Press.

Clarke, R. (2009). Dispositions, abilities to act, and free will: The new dispositionalism. *Mind, 118*(470), 323-351. doi:10.1093/mind/fzp034

Ewald, W. (1995). Comparative Jurisprudence (I): What Was It Like to Try a Rat? *University of Pennsylvania Law Review, 143*(6), 1889-2149.

Fara, M. (2008). Masked Abilities and Compatibilism. *Mind, 117*(468), 843-865. doi:10.1093/mind/fzn078

Feinberg, J. (1965). The Expressive Function of Punishment. *The Monist*, 397-423.

Fischer, J. M., & Ravizza, M. (1993). *Perspectives on moral responsibility*: Cornell University Press.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility And Control. A Theory of Moral Responsibility* Cambridge, UK: Cambridge University Press.

Frankfurt, H. (1971). Freedom of the Will and the Conept of a Person. *Journal of Philosophy, 68*, 5-20.

Fricker, M. (2016). What's the Point of Blame? A Paradigm Based Explanation. *Nous, 50*(1), 165-183. doi:10.1111/nous.12067

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin, 40*(8), 986-997.

Haukioja, J. (2013). Different notions of response-dependence. In Hoeltje, Schnieder, & Steinberg (Eds.), *Varieties of Dependence* (pp. 167-190). Munich: Philosophia Verlag.

Hieronymi, P. (2004). The Force and Fairness of Blame. *Philosophical Perspectives, 18*(1), 115-148. doi:10.1111/j.1520-8583.2004.00023.x

Hieronymi, P. (2007). Rational capacity as a condition on blame. *Philosophical Books, 48*(2), 109-123. doi:10.1111/j.1468-0149.2007.00435.x

Jackson, F., & Pettit, P. (2002). Response-dependence without tears. *Nous, 12*, 97-117.

Johnston, M. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Suppl. 63*, 139-174.

Johnston, M. (1992). How to Speak of the Colors. *Philosophical Studies, 68*, 221-263.

Johnston, M. (1998). Are Manifest Qualities Response-dependent? *Monist, 81*, 3-43.

Lacey, N. (2016). *In search of criminal responsibility: ideas, interests, and institutions*. Oxford, UK: Oxford University Press.

Mackie, J. L. (1977). *Ethics*. Harmondsworth: Penguin.

Macnamara, C. (2013). 'Screw you!', & 'thank you'. *Philosophical Studies, 165*(3), 893-914.

Manley, D., & Wasserman, R. (2007). A gradable approach to dispositions. *The Philosophical Quarterly, 57*(226), 68-75.

McGeer, V. (2012). Co-reactive attitudes and the making of moral community. In R. Langdon & C. Mackenzie (Eds.), *Emotions, imagination and moral reasoning* (pp. 299-326). New York, NY: Psycholgy Press.

McGeer, V. (2013). Civilizing Blame. In J. D. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 162-188). Oxford: Oxford University Press.

McGeer, V. (2014). Strawson's Consequentialism. In D. Shoemaker & N. Tognazzini (Eds.), *Oxford Studies in Agency and Responsibility, volume 2* (Vol. 2, pp. 64-92).

McGeer, V. (2015a). Building a better theory of responsibility. *Philosophical Studies, 172*(10), 2635-2649.

McGeer, V. (2015b). Mind-making practices: the social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations, 18*(2), 259-281. doi:10.1080/13869795.2015.1032331

McGeer, V. (2018 (forthcoming)). Intelligent Capacities: . *Proceedings of the Aristotelian Society*.
McGeer, V., & Funk, F. (2015). Are 'Optimistic'Theories of Criminal Justice Psychologically Feasible? The Probative Case of Civic Republicanism. *Criminal Law and Philosophy*, 1-22.
McGeer, V., & Pettit, P. (2015). The hard problem of responsibility. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility* (Vol. 3, pp. 160-188). Oxford: Oxford University Press.
McKenna, M. (2012). *Conversation and Responsibility*. New York, NY: Oxford University Press.
Nadelhoffer, T., Heshmati, S., Kaplan, D., & Nichols, S. (2013). Folk retributivism and the Communication Confound. *Economics and Philosophy, 29*, 235-261.
Nelkin, D. K. (2011). *Making sense of freedom and responsibility*: Oxford University Press.
Nichols, S. (2013). Brute Retributivism. In T. Nadelhoffer (Ed.), *The Future of Punishment* (pp. 25-46). Oxford: Oxford University Press.
Oshana, M. (1997). Ascriptions of Responsibility. *American Philosophical Quarterly, 34*, 81-102.
Pettit, P. (1990a). The Reality of Rule-Following. *Mind, 99*, 1-21.
Pettit, P. (1991). Realism and Response-Dependence. *Mind, 100*(4), 587-626.
Pettit, P. (1998). Noumenalism and Response-dependence. *Monist*.
Pettit, P. (2002). *Rules, Reasons, and Norms: Selected Essays*. Oxford: Oxford University Press.
Pettit, P. (2018). *The Birth of Ethics: A reconstruction of the role of nature in morality*. New York: Oxford University Press.
Pettit, P., & Smith, M. (1996). Freedom in Belief and Desire. *Journal of Philosophy, 93*, 429-449.
Pickard, H. (2013). Responsibility without blame: Philosophical reflections on clinical practice. *Oxford handbook of philosophy of psychiatry*, 1134-1154.
Ryle, G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.
Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press of Harvard University Press.
Shoemaker, D. (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics, 121*(3), 602-632. doi:10.1086/659003
Shoemaker, D. (2015). *Responsibility from the Margins*. Oxford, U.K.: Oxford University Press.
Shoemaker, D. (2017). Response-dependent responsibility; or, a funny thing happened on the way to blame. *The Philosophical Review, 126*(4), 481-527. doi:10.1215/00318108-4173422
Shoemaker, D., & Vargas, M. (2017). *Moral torch-fishing: A signalling theory of blame*. Paper presented at the APA, Pacific Division, Seattle, WA.

Smith, A. M. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics, 115*(2), 236-271. doi:10.1086/426957

Smith, A. M. (2007). On Being Responsible and Holding Responsible. *The Journal of Ethics, 11*(4), 465-484.

Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies, 138*(3), 367-392.

Smith, A. M. (2012). Attributability, Answerability, and Accountability: In Defense of a Unified Account. *Ethics, 122*(3), 575-589. doi:10.1086/664752

Smith, M. (2003). Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In S. Stroud & C. Tappolet (Eds.), *Weakness of Will and Practical Irrationality* (pp. 17-38): Oxford: Clarendon Press.

Sripada, C. (2016). Self-expression: A deep self theory of moral responsibility. *Philosophical Studies, 173*(5), 1203-1232.

Sterelny, K. (2012). *The evolved apprentice*. Cambridge, MA: MIT press.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy, 48*, 187-211.

Strawson, P. F. (1974). Freedom and Resentment *Freedom and Resentment and Other Essays* (pp. 1-25). London: Methuen.

Talbert, M. (2012). Moral competence, moral blame, and protest. *The Journal of Ethics, 16*(1), 89-109.

Vargas, M. (2013). *Building better beings: A theory of moral responsibility*. Oxford, UK: Oxford University Press.

Vihvelin, K. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics, 32*(1/2), 427-450.

Vincent, N. A. (2013). Blame, desert and compatibilist capacity: a diachronic account of moderateness in regards to reasons-responsiveness. *Philosophical Explorations, 16*(2), 178-194. doi:10.1080/13869795.2013.787443

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Wallace, R. J. (2011). Dispassionate Opprobrium: On blame and the reactive sentiments. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and Recognition: Essays on the philosophy of T.M. Scanlon* (pp. 348-372). New York: Oxford University Press.

Watson, G. (1975). Free Agency. *The Journal of Philosophy, 72*(8), 205-220.

Watson, G. (1987). Responsibility and the limits of evil: Variations on a Strawsonian theme. In F. Schoeman (Ed.), *Responsibility, character and the emotions: New essays in moral psychology* (pp. 256-286). Cambridge, UK: Cambridge University Press.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics, 24*, 227-248.

Watson, G. (2005). *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press.

Williams, B. (1995). Internal reason and the obscurity of blame. In B. Williams (Ed.), *Making sense of humanity and other philosophical papers* (pp. 35-45). Cambridge: Cambridge University Press.

Williams, G. (2017). Verantwortung, Rationalität und Urteil [Responsibility, Rationality and Judgment]. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (English version available from the author ed., pp. 365-393). Wiesbaden: Springer.

Wolf, S. (1987). Sanity and the metaphysics of responsibility. In F. Schoeman (Ed.), *Responsibility, Character and the Emotions* (pp. 64-62). Cambridge, UK: Cambridge University Press.

Wolf, S. (2011). Blame, Italian Style. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and Recognition: Essays on the philosophy of T.M. Scanlon* (pp. 332-347). New York: Oxford University Press.

(Notes inserted here by word program)

[1] Throughout this paper, I use the expression 'being responsible' as shorthand for the state or condition of *being fit to be held responsible* – i.e., having whatever it takes in a general sense to be a responsible agent. Individuals can be responsible agents in this sense without being responsible – i.e., praiseworthy / blameworthy -- *for* particular acts (or omissions) sourced in their own agency, so long as there is some suitable explanation as to why these acts do not redound to their credit/ discredit.

[2] Perhaps a similar distinction can be found in Michael McKenna's characterization of Strawsonian views, distinguishing them according to whether they are 'normative' (perhaps my 'Conventionalist View') or 'moderately realist' (perhaps my 'Indicative View') (see Chapter 2, McKenna, 2012). Certainly, there is a family resemblance here. There may be other ways of carving up this terrain, conceptually speaking. See, for instance, a recent paper by David Shoemaker (2017), whose own taxonomy of possible views differs from mine despite some suggestively similar lines of thought. To highlight one dimension of difference: his set of possible views is directed towards analyzing the property of blameworthiness or being blameworthy for some act or omission, whereas mine is directed towards analyzing the property of being fit to be held responsible, a property that blameless and blameworthy agents may share in common. That said, my so-called 'conventionalist' view is rather close to what he calls a 'dispositional response-dependent' view; though his so-called 'response-independent' view (being explicitly anti-Strawsonian in flavour) is rather different from what I call the 'indicative' view (I elaborate further on these differences in note 10 below).

[3] This view might also be termed 'Constitutivist', following Gary Watson's exegetical suggestion: "Strawson's radical claim is that these 'reactive attitudes' … are constitutive of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions" (Watson, 1987, pp., p. 257). The difficulty with Watson's remark is that it is ambiguous: is he simply saying that the proneness to reactive attitudes just is *regarding* (i.e. treating) the targets of those attitudes as responsible (this is commensurate with the Indicative View I discuss below); or is he saying (in essence) that *being* the target of such attitudes, and hence being so regarded, is all that there is to being responsible (this is commensurate with the view I am here calling 'Conventionalist')? I avoid the term 'Constitutivist' in this context because I think there is some confusion over this in the literature: some philosophers use this label for something like the 'Indicative View'; others use it for something like the 'Conventionalist View'. Indeed, some philosophers seem to go back and forth between these two views, making their own positions hard to characterize: e.g. Gary Watson, Jay Wallace, Bennett Helm, and Stephen Darwall (though I am tempted to categorize the first two as 'Indicativists' and the last two as 'Conventionalists', acknowledging that this is a matter for debate). There is a further issue that muddies the water here which I take up in footnote 7.

[4] In keeping with a Chomskyian view of things, there may be cognitive constraints shaping the grammatical norms that emerge in a given language; but this is perfectly compatible with the point I am making here.

[5] I introduce the terminology of 'response-dictated' properties, as distinguished from 'response-disclosed' properties (discussed below), to avoid certain ambiguities and confusions that have arisen in the literature with regard to so-called "response-dependent" vs. "response-independent" concepts and/or properties. This standard terminology is often taken to mark a distinction in the ontological status of the properties picked out by various concepts, as implied in Mark Johnston's original discussion (Johnston, 1989, 1992, 1998) – i.e. whether they have a 'unity' or 'integrity' that is independent of our response (Shoemaker, 2017). However, subsequent critical discussions (notably by Philip Pettit 1990a, 1991, 1998) underline a sense in which this distinction is too coarse-grained, failing to pick up a dimension of response-dependence in various concepts we use to think about the world (see too, Jackson & Pettit, 2002). For instance, in spite of there being some objective ontological unity in a particular target-property, that property may only be salient to, or significant for, us because of the way we are physically, cognitively, or emotionally structured (a point that does not emerge so clearly in Shoemaker's recent discussion). Hence, we would not have the concept of such a property absent our responses; the concept is in that sense 'response-dependently mastered', to use Jackson and Pettit's terminology, rather than 'response-dependently defined' (Jackson & Pettit, 2002). The terminology I use in this paper aims for a more pellucid way of marking these distinctions, avoiding any

further ambiguity or confusion through recycling the adjectives 'response-dependent' vs. 'response-independent'. For an excellent overview of the major cross-cutting notions of 'response-dependence' that have emerged in the literature, see (Haukioja, 2013). The relevant Pettit essays are also collected in (Pettit, 2002), which further includes a helpful overview of the main ideas.

[6] Of course, there is a distinctive, second-order way in which everyone could go wrong about the property that makes for category membership – namely, in understanding that the property in question is in fact a response-dictated property. Like the deluded fashionistas I mentioned above, people may generally believe that a particular response-dictated property is not in fact wholly determined by the norms we embrace; and it may take real philosophical work to uncover this error and/or convince people that such an error has occurred (see, for instance, Mackie, 1977 on the nature of ethical concepts). The same goes for properties that people *take* to be response-dictated, when in fact the properties they are responding to are norm-independent features of the things in question. I will discuss such a case below. My thanks to Frank Jackson for suggesting that I clarify this point.

[7] Given that the Indicative View relies on the analysis of the underlying nature of our attitudes and practices of holding responsible, one might argue that the Indicative View collapses into the Conventionalist View. However, Indicative Strawsonians would resist the idea that our attitudes and practices are simply conventional -- hence, subject to radical change or alteration. Rather these attitudes and practices are built into our very nature as human beings (e.g. like our practice of detecting and categorizing things as coloured). If the Conventionalist wants to go along with this thought, then it would be better to say that Conventionalist View collapses into the Indicative View. Perhaps this also accounts for some of the wobble I detect between Conventionalist Strawsonians and Indicative Strawsonians, with Conventionalist Strawsonians really treating the practice in question as fundamentally grounded in a non-optional form of human life.

[8] This has been variously called a 'fairness' view (e.g., Wallace, 1994) and an essentially 'communicative' view (e.g., McKenna, 2012; Watson, 2005). Of course, the Indicative View will take various forms, depending on various theorists' analyses of what normative demands are expressed/made by reactive attitudes and practices – for example, that an agent's attitudes and actions express a certain 'quality of will', or that they are appropriately 'reason responsive', etc..

[9] I follow Jackson & Pettit (2002) in using this example (see, too, papers collected in Pettit, 2002). Some may dispute whether colours are actually mind-independent objective properties of things, but for the purposes of this analogy I am clearly assuming they are. For a compelling defence of this view, see (Byrne & Hilbert, 2003).

[10] I hope this discussion makes clear why my so-called 'indicative view' is not what David Shoemaker would call a 'response-independent' view (Shoemaker, 2017). According to Shoemaker's characterization of the response-independent view, "being responsible is metaphysically prior to holding responsible" (p. 498). Hence, as he explains, advocates of the response-independent view are seemingly committed to the idea that the property of agents that makes them responsible (fit to be held responsible) is identifiable independently of "*any* reference" to our attitudes and practices of holding responsible. But my indicativist would certainly deny this, as I have been at pains to argue. Perhaps, then, the indicative view is closer to what Shoemaker calls a 'fitting-attitude' view. But this doesn't seem right either, since on this view, there is no underlying unity or integrity in the responsibility-making property – i.e. nothing that makes it *ontologically* characterizable apart from saying that agents who possess this property 'merit' (normatively speaking) a reactive response (Shoemaker, 2017, p. 508). But my indicativist would certainly deny this, perhaps adding (as I would myself) that the 'fitting attitude' view sounds either metaphysically mysterious or (at bottom) merely conventionalist.

[11] Often called 'backwards-looking' desert, but I prefer to avoid this terminology for reasons that will emerge later in this paper.

[12] One caveat: I here presume that an intuitive wedge can be driven between the pessimists' notion of substantial desert in the credit/discredit sense and their hyperbolic notion of 'ultimate' responsibility. If the pessimists' notion of substantial desert is simply *identified* with the kind of responsibility associated with an unconditioned power of choosing to do A rather than B, then of course there is no way to accomplish what Strawson sets out to do. But surely, as Strawson himself suggests, more sensible pessimists might settle for less.

[13] For a fascinating account of one such felony trial against the grain-stealing rats of Autun, see (Ewald, 1995). He outlines the wily arguments of the 16th century French defence attorney, Barthelemy Chassenée, who managed to get a judgement in favour of the rats; but it is notable that none of Chassenée's arguments questioned the presupposition that the rats were an appropriate target of criminal complaint.

[14] This objection has been forcefully articulated by Fischer & Ravizza (1993, pp., p. 18), among others.

[15] As an anonymous referee for EJP pointed out, you might regard the pessimists' desert test as a desideratum on an acceptable theory, "... but one that is permissibly abandoned if the balance of other theoretical virtues … favour … its abandonment".

[16] These are generally seen to fall into three distinct categories: (1) reason-responsive accounts (e.g., Fischer & Ravizza, 1998; Nelkin, 2011; Pettit & Smith, 1996; Vargas, 2013; Wallace, 1994; Wolf, 1987); (2) 'quality of will' accounts (e.g., McKenna, 2012; Shoemaker, 2015); and (3) 'real self' (or 'mesh') accounts (e.g., Frankfurt, 1971; Sripada, 2016; Watson, 1975). Many theorists, however, defend some kind of combination of these (for a representative recent example of why a combined approach is necessary, see: Shoemaker, 2015).Though this is an argument for another paper, I think the most plausible versions of 'quality of will' accounts (e.g., McKenna, 2012) and 'real self' accounts (e.g., Wolf, 1987) are, in effect, reason-responsive accounts. Hence, I focus on such accounts here.

[17] This kind of approach has been particularly well-developed by (Fischer & Ravizza, 1998).

[18] For critical discussion of how best to specify these counter-factual conditions, see Vargas (2013). Vargas argues, I think convincingly, that our reason-responsive capacities are more dependent on circumstance than theorists in this tradition generally admit. Although Vargas and I part company on what to make of this fact (for critical discussion, see McGeer, 2015a), his observations provide important collateral support for the view of reason-responsive capacities I go on to develop in Sections 3 and 4.

[19] In "The Hard Problem of Responsibility" (McGeer & Pettit, 2015), my co-author and I have referred to this condition as the agent's possessing the "specific" capacity to respond to the reasons before her; and not just some general capacity which she might excusably fail to exercise in her present circumstances (*pace* Wallace). The following discussion rehearses the argument we make in much greater detail in that paper (see too: McGeer, 2018 (forthcoming); Pettit, 2018, ch. 6).

[20] As a critical sidenote to the dialectic of this paper, I acknowledge that a growing contingent of philosophers resist using reason-responsive capacity as a criterion of responsibility attributions on explicitly *normative* grounds – i.e. they do not think 'fairness', in Wallace's terms, should be the 'master norm' governing when it is right or appropriate to blame people for their misdeeds (this is nicely discussed in Williams, G., 2017). Other considerations -- such as defending a victim's rights, or expressing support for moral norms, or signalling one's own refusal to countenance friendly relations with 'bad' characters – are arguably as important, if not more important, to us (see, for instance:Feinberg, 1965; Hieronymi, 2007; Scanlon, 2008; Smith, A. M., 2007; Talbert, 2012). And if that's the case, then philosophers need not be so concerned with an offending agent's putative capacity to comply with moral norms; perhaps all that matters is "a more general capacity", as Hieronymi puts it, "to stand in interpersonal relationships" (Hieronymi, 2007). One way philosophers have softened the jarring implications of this stance is to argue that blame need not involve any intent to, or interest in, sanctioning wrongdoers; indeed, it need not involve any 'reactive anger' at all (Hieronymi, 2004;

McGeer, 2013; Scanlon, 2008; but for critical pushback, see: Wallace, 2011; Wolf, 2011). Another, and for some additionally, softening manoeuvre is to follow Gary Watson (1996) in distinguishing between various type of responsibility, according to which agents may be responsible in an 'attributability' sense (their bad behaviour is a reflection of their bad character or judgement or whatever), without being responsible in an 'accountability' sense (they lack(ed) the capacity to do better). Alternatively, many philosophers now argue that, so long as agents' actions 'express' their 'judgment sensitive attitudes', they are in principle reachable by normative demands; and this is all that matters for being responsible *in an 'answerability' sense*, hence appropriately blamed for their misdeeds (Oshana, 1997; Scanlon, 2008; but see too, Shoemaker, 2011; Smith, A. M., 2005, 2007, 2008, 2012). For reasons that will become clear in Section 4, I have considerable sympathy for this answerability line of thought, though think it can be accommodated within the reason-responsive framework once we have a better understanding of the relevant capacities in hand. Hence, to my mind, it does not represent a challenge to the dominant thought otherwise questioned in this footnote – viz., that some sort of capacity for tracking and responding to normative considerations lies at the heart of any adequate (i.e. defensible) account of responsible agency. However, since I cannot defend this assumption here, I merely note that this issue is a matter of some philosophical controversy to be discussed at length elsewhere.

[21] The ideas in the remainder of this paper, but especially those in the next Section 3, are developed in greater detail in (McGeer, 2018 (forthcoming)).

[22] For recent work challenging the atemporality of standard models of capacities, see (Vincent, 2013). However, Vincent does not recognize the developmental nature of such capacities and so, to my mind, misses the significant theoretical progress compatibilists can make through shifting to a more diachronic perspective.

[23] For a defence of this view, see (McGeer, 2015b)

[24] For discussion see (Sterelny, 2012) on the importance of 'apprentice-learning' in human evolution.

[25] Indeed, one might see the Scaffolding View as simply a variant of the Indicative View, but one that has far-reaching implication for a host of issues in consequence of the skill-based account of intelligent capacities around which it's built. Some of the implications are, of course, addressed in the remainder of this paper. But one important one, not herein addressed, is that it undermines any rationale for retributive blame and/or punishment (for further discussion of this topic, see: McGeer, 2012, 2013; McGeer & Funk, 2015).

[26] Someone might reasonably wonder whether this lability in moral norms likewise affects norms governing what it takes to be a morally responsible agent, so that the nature of such

agency is itself a matter of interpersonal negotiation and development. My answer to this is 'yes'; but this is not to suggest that Conventionalists are right after all. The question remains: what is it that constrains how these norms change and develop? Is there some norm-independent feature of agents that we can see our practices as generally, if imperfectly, tracking – or not? That is the issue that divides the Conventionalist View from both the Indicative View and the Scaffolding View here defended.

[27] The term 'proleptic' comes from Bernard Williams, used specifically in reference to the attitudes and practices involved in blame (Williams, B., 1995). A proleptic view of the reactive attitudes has lately been defended by other philosophers besides myself, including, for instance, Manuel Vargas (2013) and Miranda Fricker (2016). Naturally, there are differences in the way we all develop this thought, but a family resemblance amongst the different views is clearly discernible (see, too, Williams, G., 2017 for suggestions along similar lines).

[28] This terminology comes from McGeer & Pettit (2015).

[29] Note how the proleptic view of the reactive attitudes fits very well with the enlarged list of such attitudes that Strawson originally proposed in his paper. It does not fit so well with the more restricted view of reactive attitudes (blame, indignation, resentment) defended by Jay Wallace -- again, reflecting the Indicative View that these attitudes are properly viewed as entirely backward-looking.

[30] For somewhat more detailed speculation on the role of our emotions in reactive scaffolding, see (McGeer, 2013). For the record, the argument there addresses the resistance, seen in some moral philosophers, to the 'unfortunate' emotionality of blame (e.g., Scanlon, 2008).

[31] Another issue I leave for another day is the following: we often blame people for their misdeeds even when they are dead, absent, or known to be 'hard cases' – i.e. have shown themselves to be robustly resistant to taking responsibility for their misdeeds, tending instead to offload responsibility for their actions on to others, or on to unfair/unfortunate/unhappy circumstances even, and perhaps especially, when blamed (the Robert Harris case discussed by Watson (1987) is perhaps a good example of this, though I think there may be complications here relating to the issue of acquired sociopathy that muddy a clean theoretical treatment of this case). Does the view I defend here imply that such blame is inappropriate? Or if it is appropriate, on what grounds? After all, blame can hardly play a scaffolding role in developing and/or supporting another's (skill-based) capacity for (moral) reason-responsiveness if the blame in question falls on deaf ears (in one metaphorical sense or another). And presumably our would-be blamers are in a position to know this fact about their blame in these cases. (My thanks to an anonymous referee for *EJP* for raising this worry.)

This is an important issue to address; and, though I can't go into any argumentative detail, I can at least sketch the kind of response I would give. The background account of blame (and other reactive attitudes) on which I here rely is both functional and naturalistic, appealing to how these emotions are likely to have evolved in a norm-governed cooperative species such as ours (McGeer, 2013). On this account, the attitudes we experience towards one another, especially in light of normative transgressions, have been shaped by selective pressures because of their aptness in performing the primary function of (directly) eliciting better norm-governed behaviour from conspecifics. This implies that blame is only appropriate (i.e. well-targeted) so far as it is apt for serving its primary function. And this seemingly requires two things: (1) that it's directed towards those who are *capable* of being suitably responsive to the demands being made of them; and (2) that it's directed towards those with whom the blamer is suitably connected, putting them in a position such that they can be suitably responsive to the demands being made of them.

Notice immediately that condition (1) is not a success condition: As emphasized in this section, it does not require that people actually be responsive to another's blame; only that they have the psychological wherewithal to be responsive. On my relatively modest view, that means simply having whatever it takes, psychologically speaking, to be developmentally *sensitizable* to the normative demands being made of them. For reasons I won't go into here, I maintain that the bar we should set for escaping this condition is relatively high, requiring good evidence of psychological (and ultimately neurological) disorder. So, in partial response to the challenge above, my view is that psychologically recalcitrant individuals are indeed appropriate targets of blame (modulo the presence of significant disorder). (Of course, this still leaves open the question of how proletic blame is most effectively expressed in such difficult cases; and, in this matter, I am largely in agreement with Pickard's therapeutic approach to holding responsible (Pickard, 2013), much as I disagree with her restricted understanding of (affective) blame as a punitive, essentially retributive attitude).

But, now what about the absent (including the dead)? How can it be appropriate to blame them, given condition (2) above? The line I take here is slightly more complex – and in need of careful development. But, in essence, it is this: While the primary function of blame is the scaffolding one I describe, there is nothing to rule out its acquiring additional functions ancillary to, albeit connected with, this primary function. And if that's the case, then blame may be appropriate in an extended sense so far as it serves one of more of these additional functions. For instance, commensurate with condition (1), one of the connotations of blaming someone in central cases is to mark the presence in them of a corresponding reason-responsive capacity. Hence, blame can acquire the function of signalling to others that the miscreant has (or had) the requisite capacity, perhaps even that they ought to be (have been) held accountable (blamed in the primary sense) by someone

appropriately related to them. Blame may acquire other signalling functions as well – e.g. that you yourself are prepared to adhere to, and indeed defend, norms that you take the culpable miscreant to have breached; and would hold such a person to account (blame in the primary sense) were you in a position to do so (for a rich discussion of the signalling function of blame, see Shoemaker & Vargas, 2017). The point is simply this: so far as blame acquires these additional (signalling) functions, it can be appropriate to blame in an expanded sense even when the scaffolding function cannot be directly discharged. (But note: bystanders may still be *indirectly* scaffolded in their reason-responsive capacity by way of blamers marking for them what sort of attitudes and behaviour in (non-present) others invite appropriate blame – and, hence, would invite for similar transgression in their case as well.)

[32] This objection, though familiar enough from the classic debates (e.g., as between Ayer (1980) and Chisholm (1982)), finds recent and powerful articulation in (Clarke, 2009).

[33] Again, see note 29 for my response to the challenge that we blame others even in situations where a practical concern of this sort seems less relevant to 'appropriate' blame.

[34] In McGeer & Pettit (2015), we call such exhortative claims, 'evocatives', where we also provide a more in-depth discussion of this general phenomena, as well as a more detailed analysis of 'you could have done otherwise'. See too, McGeer (2018 (forthcoming)) and Pettit (2018, Ch. 6).

[35] There has been much debate in the literature as to whether the ordinary notion of desert has an entirely backward-looking cast, or whether it has a forward-looking dimension to it as well. Philosophers divide on this question, including even those who use empirical research into "folk moral psychology" to replace, or at least supplement, the standard philosophical appeal to "intuition". Thus, some 'Ex-phiers' take certain studies to show that human beings have a 'brutely retributive' (i.e. purely backward-looking) moral psychology (Nadelhoffer, Heshmati, Kaplan, & Nichols, 2013; Nichols, 2013); whereas others present results demonstrating a distinctly forward-looking interest/ concern (Funk, McGeer, & Gollwitzer, 2014; McGeer & Funk, 2015). Obviously, this remains a matter of some controversy. So, here I make another suggestion: that the problem of desert is not well conceptualized in terms of a backward-forward looking divide (perhaps partially explaining these contradictory empirical results); and that, as theorists, we would be well-advised to explore a different question. To wit: how are folk judgements of desert affected by the way people regard others' (intentional) actions – as one-off 'out of character' doings, or as doings that indicate (what could be) a larger pattern of behaviour if steps are not taken to ensure otherwise? We already know from numerous psychological studies that ordinary folk are very ready to see individual actions as attributable to 'underlying character' (versus situational factors). Here I suggest that this pre-occupation with

'underlying character' is, in fact, better understood as a pre-occupation with locating an individual's actions in the larger framework of her potential doings that we think she can be brought to achieve, if only by our encouragement and insistence. And this in turn points to a natural reframing of the desert question along the lines I suggest here.

[36] Earlier versions of this paper were presented at various workshops and colloquia, and I am grateful for the many helpful questions and comments I received on those occasions. These include an initial foray at a workshop on reactive attitudes at the Universität Duisberg-Essen in June 2016, followed by colloquia presentations at the University of Melbourne (March 2017), the University of Sydney (April 2017), the University of British Columbia (April 2017), the Australian National University (May 2017), and the UCHV Fellows Seminar at Princeton University (November 2017). Impossible to catalogue the many ways in which the feedback I received improved the quality of this paper; but thanks are owed especially to Daphne Brandenburg, Susan Brison, David Hilbert, Frank Jackson, Karen Jones, Jeanette Kennett, Philip Pettit, Francois Schroeter, Laura Schroeter, Michael Smith, Daniel Stoljar, Monique Wonderly, and an anonymous referee for the *European Journal of Philosophy*. This work was supported by the Australian Research Council [grant number DP140102468] and generous ongoing support from Princeton University.