

## **HOLDING RESPONSIBLE RECONSIDERED**

Larisa Svirsky

Following Strawson, many philosophers have claimed that holding someone responsible necessitates its being appropriate to feel or express the negative reactive attitudes (e.g., resentment) toward her. This view, while compelling, is unable to capture the full range of cases in which we hold others responsible in ordinary life. Consider the parent who holds her five-year-old responsible for not teasing his sister, or the therapist who holds her patient responsible for avoiding self-injurious behavior. Holding responsible in such cases requires enforcing normative expectations, but these norms can (and typically should) be enforced without involving the negative reactive attitudes. To demonstrate this, I consider how responsibility attributions function in psychotherapy, as well as in other contexts where the negative reactive attitudes do not have a natural home.

### **1. Introduction**

An undertheorized aspect of our responsibility practices is that we routinely hold agents responsible even when they are impaired or immature (i.e., so-called “marginal agents”).<sup>1</sup> This holding responsible is an essential part of parenting, teaching, and psychotherapy, for example. A parent who chose not to hold his daughter responsible for anything merely because she was a child, or a therapist who chose not to hold her patients responsible for coming to the majority of their appointments as scheduled, would seem misguided and even negligent. Yet standard philosophical approaches to responsibility often identify being responsible with the possession of certain psychological capacities, such as rationality and self-control, in which marginal agents are deficient. Can these theories of responsibility be reconciled with our responsibility practices? If not, how should we understand what the parent and the therapist take themselves to be doing?

In general, the distinction between being responsible and holding responsible has proved helpful for untangling various judgments about responsibility. Roughly, to be responsible is to satisfy whatever conditions are necessary for responsible agency; one holds someone responsible for an action or attitude when one holds her to account for it. In “On Being Responsible and Holding Responsible,” Angela Smith characterizes this distinction as follows: “...to say that a person is morally responsible for some thing is to say that it can be attributed to her in the way that is required

in order for it to be a basis for moral appraisal.” In contrast, ‘holding responsible’ refers to “something not about the relation between an agent and her actions or attitudes, but about *our* relation to the agent and our (usually negative) assessment of her conduct”.<sup>2</sup>

Much of the literature about holding responsible concerns whether being responsible or holding responsible is prior (either explanatorily or metaphysically speaking).<sup>3</sup> My primary focus in this paper, however, lies elsewhere. In “Freedom and Resentment,” Strawson emphasizes the responses we are inclined to have toward people we regard as responsible agents (“the range of reactive feelings and attitudes which belong to involvement or participation with others in interpersonal human relationships...”) as fundamental to understanding the nature of responsibility.<sup>4</sup> Following Strawson, many philosophers have claimed that holding someone responsible necessitates its being appropriate to feel or express the negative reactive attitudes (e.g., resentment, indignation) toward him.<sup>5</sup> This view, while compelling, is ultimately unable to capture the full range of cases in which we hold others responsible in ordinary life.<sup>6</sup> Consider the parent who holds her five-year-old responsible for not teasing his sister, or the therapist who holds her patient responsible for avoiding self-injurious behavior. Paradigmatically, these instances of holding responsible involve enforcing normative expectations, but an appreciation of the agent in question’s immaturity or impairment leads this enforcement not to involve the negative reactive attitudes in any way. These cases also suggest the importance of another aspect of Strawson’s project, namely that our responsibility practices are bound up with our participation in interpersonal relationships. These relationships are an important source of normative expectations, and therefore the site of a special kind of authority to hold one another responsible.<sup>7</sup>

My aim here is to offer a characterization of holding responsible that allows us to appreciate and understand the wide range of forms this notion takes in practice. Before arriving at such a view, I will first describe some alternatives. I will begin with Strawson’s characterization of responsibility,

focusing especially on the connection between responsibility and the reactive attitudes on his view. I will then consider an argument from R. Jay Wallace that responsibility is only connected to some of the things Strawson characterizes as reactive attitudes (in particular, the negative reactive attitudes – that is, resentment, indignation, and guilt, and not e.g., gratitude). Wallace argues that this connection is forged via the notion of normative demands and expectations; agents are responsible when they can be sensitive to these expectations, and violating them justifies another's feeling or expressing the negative reactive attitudes.<sup>8</sup> I want to endorse Wallace's suggestion that responsibility is essentially related to normative expectations but reject the connection he draws between holding responsible and the justifiability of the negative reactive attitudes. To demonstrate this point, I consider how responsibility attributions function in psychotherapy (and discuss my view on this subject in relation to Hanna Pickard's), as well as in other contexts where it does not seem that the negative reactive attitudes have a natural home.

With these considerations in place, I then lay out my positive view of holding responsible as essentially involving the enforcement of norms, where norm enforcement takes on a much wider variety of forms than is often appreciated.<sup>9</sup> In particular, I argue that one can enforce a norm by praising someone's having met it or reminding someone of its existence in a forward-looking way, as well as by responding to norm violations; these other responses are just as central to holding responsible as blame is. This broader theoretical conception of holding responsible affords us a better understanding of our responsibility practices, in which marginal agents are held responsible despite often being exempted from some range of the reactive attitudes. Moreover, this view of holding responsible makes room for a more expansive characterization of who counts as a responsible agent – that is, for what it means to be responsible. Ultimately, I claim that cases involving these agents show us how being and being appropriately held responsible are inextricably linked with normative expectations, which importantly include those that arise in our interpersonal

relationships. Though I came to this view of responsibility in an attempt to do justice to cases involving marginal agents, I will go on to argue that it in fact applies to all of us. Nevertheless, it is of particular importance for understanding how our responsibility practices involve members of vulnerable populations.

## **2. Responsibility and the Reactive Attitudes**

In “Freedom and Resentment,” Strawson considers whether determinism, if true, undermines the legitimacy of our responsibility practices. He argues that the best way to resolve this question is to consider the character of these practices (including how we respond to people we regard as responsible agents, and when we excuse or exempt people from being held responsible in everyday life) and to see if believing the truth of determinism rationally requires us to opt out of them. Strawson suggests that regarding someone as a responsible agent disposes us to have what he calls the reactive attitudes toward them, which are “...essentially reactions to the quality of others’ wills toward us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern”.<sup>10</sup> Resentment, for example, is “...a reaction to injury or indifference”.<sup>11</sup>

Strawson contrasts the reactive attitudes with the objective attitude, which we characteristically take up with those we cannot properly regard as responsible agents due to impairment or immaturity.<sup>12</sup> As he writes,

“The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in interpersonal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude toward someone is wholly objective, then

though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.”<sup>13</sup>

The fundamental connection between responsibility and the reactive attitudes that Strawson urges us to recognize is not only logical, but also psychological; when we regard others as capable of responsible agency, and as expressing something about the quality of their wills in their behavior, we are disposed to respond with the reactive attitudes in light of what they have done. This orientation toward others is necessary for one to be able to engage in the kind of intimacy that Strawson says defines interpersonal human relationships, the relationships that we have with other participants in the moral community. When we are led to believe that others are not able to participate in this kind of relationship, our suspension of the reactive attitudes is a natural consequence of the shift in the way that we see them. We are, when taking up the objective attitude, no longer prone to viewing that person as a fellow community member, but rather “to see[ing] him, perhaps, as an object of social policy; as a subject for what, in a wide range of senses, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided...”<sup>14, 15</sup>

Strawson describes a set of conditions that make us inclined to excuse or exempt people from the reactive attitudes, including ignorance, coercion, immaturity, or psychological abnormality. These conditions are meant to be familiar to us from ordinary life as reasons to abstain from responding as we normally would with someone we consider a responsible agent. Finding out that an agent was ignorant or coerced casts doubt on whether she has genuinely manifested the quality of will we initially took her to have (e.g., good will or indifference). An agent’s immaturity and psychological abnormality, on the other hand, may present obstacles to her participation in interpersonal relationships such that we are no longer prone to feel the reactive attitudes toward

her.<sup>16</sup> These claims of Strawson's are intended to be justified by reflection on our existing responsibility practices, and not by independent theoretical investigation of what conditions are required for responsible agency. Gary Watson elaborates on this point when he says, "Holding responsible is as natural and primitive in human life as friendship and animosity, sympathy and antipathy. It rests on needs and concerns that are not so much to be justified as acknowledged."<sup>17</sup>

In *Responsibility and the Moral Sentiments*, R. Jay Wallace takes himself to be following in the Strawsonian tradition of considering our responsibility practices as the ground for understanding what responsibility is and could be. Ultimately, he argues that being a responsible agent is fundamentally a matter of possessing normative competence (i.e., in Wallace's terminology, the powers of reflective self-control); in particular, it is a matter of possessing "the power to grasp moral reasons and the power to control...[one's]...behavior in accordance with them".<sup>18</sup> But he arrives at this claim by considering when it would be fair to hold individuals responsible in the sense of their being appropriate objects of the reactive attitudes, where the relevant norms of fairness are supposed to be drawn from everyday life.<sup>19</sup> For example, Wallace claims that these norms of fairness dictate that resentment is appropriate just when someone is capable of understanding moral reasons and capable of guiding her actions to be in accord with them, but has failed to do so without apparent excuse.

Wallace also holds a distinct view about the connection between responsibility and the reactive attitudes. In particular, he claims that Strawson's characterization of the reactive attitudes is much too broad. In support of this claim, he argues that Strawson and his commentators tend to elide two distinctions: "...the distinction between reactive attitudes and other attitudes one might take toward persons..." on the one hand, and the distinction between the range of attitudes that are associated with the participant stance versus the objective stance on the other.<sup>20</sup> Though certain kinds of love, forgiveness, and gratitude may only be felt toward those with whom we take up the

participant stance, it does not follow that those attitudes bear the same kind of relationship to responsibility as resentment, indignation, and guilt.

Wallace insists that these latter attitudes, unlike the others that Strawson calls reactive attitudes, derive their distinctive role in our responsibility practices from their connection to normative expectations. As he writes, “Resentment, indignation, and guilt are essentially tied to expectations that we hold ourselves and others to; susceptibility to these emotions is what constitutes holding someone to an expectation. This mutual dependence of emotion and expectation distinguishes the reactive attitudes...”<sup>21</sup> For Wallace, possession of the capacities required to be properly held to normative expectations is constitutive of responsible agency, and resentment, indignation, and guilt (i.e., the negative reactive attitudes) are fitting responses to violations of such expectations. “To hold someone to an expectation,” Wallace writes, “is to be susceptible to a certain range of emotions (i.e., the negative reactive attitudes) if the expectation is violated or to believe it would be appropriate for one to feel those emotions if the expectation is violated.”<sup>22</sup>

But Wallace and Strawson are both committed to considering our theories of responsibility as answerable to the ways the concept of responsibility is understood and instantiated in ordinary life. If we take this commitment seriously, we must recognize that there are individuals who participate in our responsibility practices, and who satisfy Wallace’s condition of being properly subject to normative expectations, who nonetheless often are and should be shielded from the negative reactive attitudes. I will sketch what I mean by this briefly here and return to these considerations later on. Consider, for example, how parents hold young children responsible for sharing their toys or behaving themselves in public, how an alcoholic’s therapist might hold them responsible for a relapse after a period of sobriety, or how an employer might hold an employee responsible for completing the tasks associated with their job. While holding responsible may take a

number of different forms in these situations, they need not be seen as situations where resentment or indignation could be appropriate.

What unites these cases is the fact that there are norms that hold between the parties in question and holding responsible involves the enforcement of those norms. In other words, we ought to accept Wallace's claim that holding responsible appears to be essentially connected to normative expectations. We should resist his suggestion, however, that enforcing those expectations necessarily involves the appropriateness of the negative reactive attitudes. This is not to deny that we are often prone to feel these attitudes toward those we hold responsible, or that those attitudes are responses to our sense of the other person as an agent. It is rather to question whether holding someone responsible means that it *must* be appropriate to feel the negative reactive attitudes toward her. If holding responsible is, as Watson suggests, a means "of expressing our concerns and demands about our treatment of one another," it is only natural for that expression to take on emotional significance.<sup>23</sup> But those concerns about our treatment of others cut both ways; though a parent might be irritated by his child's failure to share, or a therapist might be disappointed at her patient's relapse, they should (and in many cases do) also recognize the vulnerability of the parties in question, and the difficulties these parties may have in trying to satisfy the expectations placed on them. This recognition in turn mitigates the propensity the parent or therapist might have to find it appropriate to feel, much less express, the negative reactive attitudes.

### **3. Responsibility in Psychotherapy**

In order to make this claim more vivid, let us consider the role of responsibility in therapeutic relationships. Finding a sense of 'holding responsible' that is appropriate to therapeutic relationships requires steering between the Scylla and Charybdis of underestimating the agency of people with mental illness on the one hand and viewing them as fitting objects of their therapists'

affective blame on the other.<sup>24</sup> On the one hand, it is common to hear statements such as the following: “Like diabetes and Alzheimer’s, depression is an illness that befalls you. You are not responsible...or to blame [for that]. Your only responsibility is to get treatment and, once this has begun, to stay with the therapy until you are well again.”<sup>25</sup> On the other hand, clinicians do in some cases blame their patients for their symptoms. Richard Friedman, a psychiatrist, gives one such example involving a patient with chronic depression who had been referred to him by another provider,

“...a man in his early 30s, who had suffered from depression since his teenage years. In six years of psychotherapy, he had been given nearly every antidepressant under the sun, but his mood hadn’t budged. Weeping in my office one day, he explained that he was depressed because he was a failure and a whiner. ‘Even my therapist agreed with me,’ he said. ‘She said that maybe I don’t want to get better.’”<sup>26</sup>

Friedman reports that, “About a month later, I saw this patient respond remarkably well to a novel treatment. Free of depression at last, he was joyful and relieved — an odd reaction, you must admit, from someone who secretly wished to be ill.”<sup>27</sup>

Friedman suggests that this impulse to blame patients for their symptoms can arise when therapists are frustrated with their behavior or lack of response to treatment, which can create serious obstacles to a therapeutic alliance. Consider, for example, this description of patients with personality disorders cited by Hanna Pickard, a social worker and philosopher: “In his landmark study of staff’s attitudes to service users with personality disorder in three High Security Hospitals in the UK, Len Bowers suggests:

“The generally hopeless, pessimistic attitudes of carers can be seen to originate in the difficult behaviors of personality disorder patients. They bully, con, capitalize, divide, condition, and corrupt those around them. They make complaints over inconsequential or nonexistent issues

in order to manipulate staff. They can be seriously violent over unpredictable and objectively trivial events, or may harm and disfigure themselves in ways that have an intense emotional impact on staff. If this were not enough, they also behave in the same ways toward each other, provoking serious problems that the staff have to manage and contain.”<sup>28</sup>

Pickard considers this description “in many ways accurate,” and she argues that it is vital for therapists to hold their patients with personality disorders responsible precisely because these conditions are “disorders of agency”.<sup>29</sup> By this Pickard means, “Core diagnostic symptoms or maintaining factors of...[these] disorders...are actions and omissions: patterns of behavior central to the nature or maintenance of the condition. For instance, borderline personality disorder is diagnosed in part via deliberate self-harm and attempted suicide, reckless and impulsive behavior, substance use, violence, and outbursts of anger...”<sup>30</sup> Pickard suggests that it is a mistake simply to see these individuals as suffering from diseases that befall them and thereby render them wholly incapable of being or being appropriately held responsible. But though she emphasizes the importance of responsibility in the treatment of personality disorders, she cautions against regarding these agents (and those in psychiatric treatment more generally) as appropriately subject to the full range of responsibility responses, including the negative reactive attitudes.

Her proposal is that the appropriate stance for clinicians to take toward their patients involves holding them responsible but not blaming them for their behavior, even when that behavior is quite harmful to themselves and others, because blame is detrimental to the patient and to the therapeutic relationship. As she writes, “Blaming service users [i.e., psychiatric patients] may trigger feelings of rejection, anger, and self-blame, which bring heightened risk of disengagement from treatment, distrust and breach of the therapeutic alliance, relapse, and, with service users with personality disorder, potentially even self-harm or attempts at suicide: it is essential that compassion and empathy be maintained.”<sup>31</sup>

How exactly are we to make sense of the claim that patients with personality disorders are responsible for doing very bad things but are not to blame for them? In order to make out this position, Pickard insists that there is a familiar sense of ‘responsibility’ that is not moral or connected to the negative reactive attitudes. She describes this sense as follows:

“Effective clinical treatment presupposes that service users are responsible for their behavior insofar as they have conscious knowledge of what they are doing, and can exercise choice and at least a degree of control over the behavior... [T]his is a traditional and commonsense idea about what it means to be responsible, applicable not only to service users, but to us all. This idea of responsibility is essentially linked, not to morality and the reactive attitudes, but to agency. Crucially, on this view, we are responsible for all our actions, whether or not they are right, wrong, or neutral from a moral point of view. We are responsible for our actions because we are their agents: insofar as we know what we are doing, and can exercise choice and control our behavior, what we do is up to us.”<sup>32</sup>

Pickard argues that the sense of responsibility relevant to clinical contexts is non-moral because, “behaviors like self-harm, substance abuse, and obsessive rituals, can be damaging to the person without necessarily damaging others”.<sup>33</sup> On her view, much of the behavior that clinicians should hold their patients responsible for is harmful, but it is not morally bad because it only or primarily affects themselves; she also suggests that such behavior is somewhat understandable as an expression of or method of coping with emotional distress.

In addition, Pickard claims that if clinicians regarded their patients as morally at fault for their behavior, they would regard themselves as entitled to subject them to the negative reactive attitudes, which would undermine the effectiveness of their treatment. In some cases, Pickard argues that clinicians may appropriately form detached judgments that their patients are blameworthy when they cause harm to themselves or others without excuse, but this blame lacks the characteristic

‘sting’ of affective blame and so does not interfere with promoting that patient’s recovery. To clarify what she means by this ‘detached’ sense of blame, she considers how it occurs in parent-child relationships.

“For instance,” she writes, “good parenting routinely involves pointing out when a child has shown disregard or ill will toward a sibling, and indeed imposing negative consequences for it. That is part of bringing up children to treat others, including rivals, with regard and respect. Sometimes, no doubt, parents do this in such a way that the child feels bad and blamed. But a loving parent can often help a child understand that their behavior toward a sibling is neither decent nor permitted, without the child feeling ‘stung’.”<sup>34</sup>

In general, detached blame seems to involve a cognitive judgment that the person in question has caused harm, and an imposition of negative consequences, where those consequences are not intended to be of an emotional nature.

I am broadly sympathetic to Pickard’s view about how it is appropriate to hold patients responsible without subjecting them to negative reactive attitudes in the context of therapeutic relationships. I believe, however, that we can (and should) understand the significance of this without invoking a different, non-moral sense of responsibility. Pickard’s argument that the sense of responsibility applicable to service users is non-moral depends on two assumptions about behaviors that are partially constitutive of disorders of agency: first, that they only or primarily cause harm to self rather than others, and second, that this means that they are not morally bad. Both of these claims are controversial and stand in need of further defense.

More importantly, by invoking a non-moral sense of responsibility to understand what responsibility means in clinical contexts, Pickard leaves open the possibility that there is an important discontinuity between service users and the general adult population with respect to what it means for them to be responsible. This seems antithetical to her overall aim to humanize these

agents, as exemplified in the following passage: “As Angela Smith elegantly points out: ‘being held responsible is as much a privilege as it is a burden. It signals that we are a full participant in the moral community’.<sup>35</sup> In other words, in holding service users with disorders of agency responsible, we treat them as *one of us*—as belonging with us, as equals.”<sup>36</sup>

Happily, one can understand the sense of responsibility relevant to clinical contexts as moral without that entailing the appropriateness of the negative reactive attitudes. Indeed, the example Pickard gives of detached blame in a parent-child relationship still seems to be moral blame; the child in question has “shown disregard or ill will toward a sibling,” though the kind of blame that is warranted avoids the characteristic ‘sting’ associated with the negative reactive attitudes.<sup>37</sup> In some cases that arise in clinical contexts, it seems fitting to recognize that psychiatric patients are responsible for wrongful actions they perform but excuse them from affective blame on the grounds that they do not deserve it given the nature of the difficulties that they face. In other cases, where the individual in question doesn’t have an excuse, one might still think that therapists are obliged (in light of professional norms, and perhaps other considerations) to exhibit a degree of emotional detachment toward their patients that is incompatible with affective blame. Thus, holding responsible may entirely avoid the negative reactive attitudes, even if one is still considering moral responsibility.<sup>38</sup>

Another feature of Pickard’s view that is worth mentioning is her emphasis on the instrumental goods associated with the stance of responsibility without affective blame in clinical contexts. That is, she argues that holding responsible is beneficial, and the characteristic ‘sting’ associated with the negative reactive attitudes is damaging, to patients and to therapeutic relationships. These claims strike me as intuitive and important, and I have no interest in disputing them. But I also want to bring out something Pickard touches on but does not develop in much detail, namely that recognizing someone’s agency also constitutes a way of showing her respect; this

undoubtedly influences the quality of a relationship, therapeutic or otherwise. Such a recognition is fundamentally worthwhile not just because it promotes good behavior or psychological health, but because it is a part of treating the individual in question as a person.

Arguably, part of what makes therapy effective (when it is) is that therapeutic relationships are intended to serve as sources of norms, whose content and conditions of enforcement are fitting given the aims of treatment. A good therapist will recognize her patients' strengths and abilities alongside their vulnerabilities and impairments, and therefore recognize that both therapist and patient have a role to play in determining what those aims should be and how they might be achieved. This collaborative work means not only that therapists are better positioned to enforce norms with their patients in an understanding and compassionate way, but that therapists and patients have a special authority to hold each other responsible that stems from the norms created in their relationship. For example, if a patient has pledged to her therapist to abstain from self-injurious behavior for a length of time and falls short of this expectation, her therapist has grounds to hold her responsible that strangers and acquaintances, and perhaps even some people in close relationships with her lack. This holding responsible, again, need not involve the negative reactive attitudes or other harsh treatment, but might involve a reminder of the relevant expectations or an investigation of what the patient can do to help herself uphold them going forward.<sup>39</sup> These responses reflect an appreciation of the difficulties associated with recovery, but also the role of the patient's own agency in making recovery possible.

#### **4. The Core of Holding Responsible**

What does it mean to treat someone as a person by recognizing her agency? On Strawson's view, we do this when we take up the participant stance toward her, when we see her, "as a morally responsible agent, as a term of moral relationships, as a member of the moral community".<sup>40</sup> This

way of regarding another person, he claims, creates a propensity to feel the reactive attitudes, which he characterizes as, “essentially natural human reactions to the good or ill will or indifference of others toward us, as displayed in their attitudes and actions”.<sup>41</sup> Though we can suspend the reactive attitudes temporarily even with those whom we regard as responsible agents in order to “[seek] refuge from the strains of involvement,” Strawson argues that this could not be our normal mode of responding to these agents.<sup>42</sup>

It is unclear whether, on Strawson’s view, one could suspend some range of the reactive attitudes while nevertheless having others. For example, it is unclear whether he views it as psychologically possible to feel emotions like gratitude or admiration at a person’s good will without also being prone to feel resentment or indignation at her ill will or indifference. Thus, Strawson may or may not be willing to say that we are able to take up the participant stance toward those whom we also regard as properly shielded from the negative reactive attitudes. Wallace, on the other hand, separates the participant stance (i.e., the stance we take up toward fellow persons) from holding responsible. He argues that only the latter is to be identified with either feeling the negative reactive attitudes in light of a normative expectation’s being violated or judging it appropriate to feel those attitudes in such circumstances.

Rather than responding directly to Strawson’s and Wallace’s views on this subject, I would first like to take a step back and reflect on the variety of ways in which we respond in ordinary life to those we regard “as a term of moral relationships,” and what unifies those responses. After considering several cases meant to be representative of this variety, I will develop a distinct positive characterization of what it means to hold responsible, which I take to be connected to treating someone as a person.<sup>43</sup> Ultimately, I want to accept Wallace’s view that to hold someone responsible is to hold her to normative expectations, but to question whether this bears any obvious connection to the negative reactive attitudes. In clinical contexts like the ones Pickard describes, for example,

there seem to be genuine instances of holding responsible where the parties in question don't feel, or judge it appropriate to feel, those attitudes. Moreover, we should not think this is unique to therapeutic relationships. I will begin by discussing a therapeutic case, before describing other examples that also support this view of holding responsible.

Consider how a therapist might respond to an addict who has relapsed after a period of sobriety. Though it is understandable that she might feel some degree of disappointment at her patient's behavior, it seems inappropriate for her to meet this situation with anger or resentment, or to regard herself as justified in doing so. Recovering from an addiction is extremely difficult, and relapse is a common part of this process; fully appreciating this seems incompatible with viewing anger as an appropriate response to this patient. Nonetheless, I claim (with Pickard) that it is appropriate for her to hold her patient responsible. It is a common experience for addicts who relapse to feel that they are back at square one with respect to their recovery, and to question whether they have it in them to get better. By counseling her patient not to discount the value of his previous abstinence and letting him know that she still expects him to move forward in his recovery, she reminds him of norms that she still regards him as capable of upholding despite his relapse.

Consider, too, how a father might respond to a four-year-old child who refuses to share her toys with her sister. Again, it seems inappropriate for him to respond by feeling or expressing the negative reactive attitudes, or thinking they are justified in this situation. After all, it is typical for young children to need to learn they ought to share, and it would be excessively harsh to respond to normal development with resentment or anger. But it does seem appropriate for the father to hold his daughter responsible, which might involve explaining the value of sharing, or asking his daughter to imagine what it would be like if no one shared with her, as well as the imposition of consequences (e.g., taking the toy away temporarily). In doing so, the father enforces a norm that his daughter has violated (e.g., "You have to share with your sister because that is fair, and if you don't, I will take the

toy away”), but without the kind of cognitive and emotional responses Wallace treats as constitutive of holding someone responsible.<sup>44</sup>

These cases share some important features. First, they are situations where it seems fitting for the therapist and parent respectively to hold responsible without this involving the negative reactive attitudes in any way. Second, this holding responsible involves the enforcement of norms that have previously been violated by the parties in question. Third, these cases involve marginal agents – that is, individuals who are deficient in the cognitive and volitional capacities often associated with responsible agency. One might assume that these deficiencies alone serve to explain why the addict and the child are not fitting targets of the negative reactive attitudes. According to this line of thought, though it may be appropriate to act as if we hold them responsible, they are not responsible in the same sense or to the same degree as psychologically normal adults. Thus, one might worry that it is illicit to infer from these cases that holding responsible properly understood does not require feeling or regarding oneself as justified in feeling the negative reactive attitudes. I do not share this view about the responsibility of marginal agents. As I see it, marginal agency is not something that always excuses one from blame for norm violations, but rather something that shapes the norms one is held to and tone of norm enforcement. But more immediately, I also believe there are examples of holding non-marginal agents responsible that do not involve the presence or justifiability of the negative reactive attitudes. Let us consider one below.

Imagine that you are an employer and that one of your employees has recently lost his mother. His grief is interfering with his doing his job effectively, and you need him to complete one particular project or he will lose an important client. Although this is a delicate situation, it seems appropriate for you to hold him responsible for completing the tasks associated with his job. You might do this by reminding him that he knows this project better than anyone else in the office, and though you know he is having a difficult time, you expect him to be able to meet his deadlines; if he

can't, you will have to give the account to someone else. But assuming he falls short of these expectations, it seems excessively harsh to see anger or resentment as a fitting response. Here again, holding responsible involves the enforcement of a norm – in this case, a norm with respect to his job performance, but does not involve the justifiability of the negative reactive attitudes.<sup>45</sup>

In this case, as well as the two I considered above, the negative reactive attitudes are an inapt response, and a willingness to feel those attitudes or see them as justified seems to evince an impatience with or lack of compassion for the person in question. But the responses that the parent, therapist, and employer in fact have are nevertheless instances of holding responsible. They involve the enforcement of norms, and the recognition of the person in question as an agent – in particular, one who is capable of appreciating and upholding the relevant expectations. This norm enforcement that lies at the core of holding responsible is not merely a way of incentivizing good behavior, but of showing the other party a kind of respect. It reflects a presumption of reciprocity between the person holding responsible and the person being held responsible – that you can expect something of each other, and are able to appreciate the grounds for those expectations in your relationship.

Notice that if this is how we understand what it means to hold someone responsible, we do not only do this when we regard her as having violated the relevant norm. We also hold her responsible when we praise or give credit to her for doing what she was supposed to, or draw her attention to the relevant norm in a forward-looking way.<sup>46</sup> As children get older and are told that they are entrusted with more responsibilities by their parents (say, to walk themselves home from school safely and without too many diversions), they are thereby being held responsible on this view. Similarly, one is held responsible when one takes on a new job and is told on the first day what one's responsibilities are – that is, what others expect of one in this role. Or suppose someone sets a goal of abstaining from drinking for a certain length of time and succeeds in doing so. Her therapist's

praising her for this constitutes an instance of holding responsible (provided she would also have been held responsible if she had not met the relevant expectations).

The idea that the sorts of cases described in the paragraph above are instances of holding responsible lends further support to the idea that holding responsible does not necessitate the appropriateness of the negative reactive attitudes. For these are clearly cases where the negative reactive attitudes are not called for. Though they involve normative demands, those demands have either been satisfied, or the person in question has not had a chance to satisfy them yet. In addition, I have described cases here that involve both marginal and non-marginal agents, and the sense in which they are held responsible is fundamentally the same.

## **5. Conclusion**

In summary, I want to highlight four aspects of my view of holding responsible. First, on this view, the core of holding someone responsible is regarding her as accountable for upholding a norm. Second, we do this not only when norms have been violated, but also when they have been upheld, or when we want to remind someone of that norm in a forward-looking way. Third, holding someone responsible does not thereby mean it is appropriate to feel, or regard oneself as justified in feeling, the negative reactive attitudes. Finally, this conception of holding responsible applies to marginal and non-marginal agents alike.

I began developing this view about the nature of holding responsible in order to do justice to how we hold marginal agents responsible, which frequently has nothing obviously to do with the negative reactive attitudes. The view I arrived at shares elements with both Wallace's and Pickard's views of holding responsible. In particular, I think Wallace is right to emphasize the connection between holding responsible and normative demands, although I disagree that to hold someone responsible, one must judge that they have fallen short of such demands, or that they are thereby the

appropriate object of one's resentment or indignation. Likewise, I endorse Pickard's suggestions that one can hold someone responsible while nonetheless seeing it as appropriate to protect them from the characteristic 'sting' associated with the negative reactive attitudes, and that this is an important part of therapeutic relationships. Unlike Pickard, however, I claim that the sense of responsibility that is operative in clinical contexts is garden-variety accountability.

The motivation I have given for my positive view of holding responsible primarily comes from the examples I have given, which represent some important and undertheorized ways we use this concept in ordinary life. These cases, I think, demonstrate the value of understanding holding responsible as broader than seeing someone as the fitting object of the negative reactive attitudes, or even as deserving of blame. But one might worry that I have not given sufficient argument that my view describes the core of holding responsible. In particular, one might question whether the examples I have given are ones where the therapist, parent, and employer genuinely hold the relevant party responsible, as opposed to acting *as if* they do for pragmatic reasons. I see no reason to describe these cases as instances of pretense; they seem to involve perfectly reasonable responses to another person in light of regarding her as subject to a norm. The only motivation for resisting the claim that these cases are instances of holding responsible seems to come from including the justifiability of the negative reactive attitudes in the definition of holding responsible. Such a definition would beg the question against my view.

The concept of responsibility surely functions in a number of ways in ordinary life and philosophical theorizing. Nonetheless, I take myself to have identified a central strain of responsibility understood in terms of accountability, which is essential to correctly understanding what is going on in the cases I have considered.<sup>47</sup> In other words, this conception of holding responsible is an important one, even if it is not exhaustive. In particular, by rejecting the idea that holding someone responsible means thereby finding it appropriate to be angry with her, my view

makes it clear that we do in fact hold marginal agents responsible. Moreover, careful attention to these practices shows us how to understand Strawson's powerful suggestion that responsibility bears an essential connection to interpersonal relationships even if holding responsible is broadened beyond cases that involve the negative reactive attitudes. Relationships are sources of norms, and make it possible to adjust and enforce expectations in accordance with our understanding of each other.

All of this allows us to begin to see that when we hold young children, for example, responsible, we are not merely pretending that they are responsible in order to teach them how to behave. Rather, we are engaging in the very same kind of practice that we engage in with non-marginal agents, namely enforcing norms to which the agent in question is subject. Recognizing this same underlying structure in our responsibility responses to marginal and non-marginal agents in turn allows us to appreciate the continuity between these agents as participants in the moral community.

*The Ohio State University*

---

## NOTES

<sup>1</sup> By responsibility practices, I mean to refer to our evaluations of people as responsible or not, and the ways in which we hold people responsible in light of their actions and attitudes. Examples of such practices include blaming, punishing, crediting, demanding an apology, and reminding someone what we expect of them going forward. I use the term "marginal agent" following David Shoemaker, who characterizes marginal agency as occurring "at the boundaries of our interpersonal community, where agents tend to strike us as eligible for some responsibility responses but not others" (Shoemaker (2015) 4). This category includes a wide variety of agents, such as children, addicts, adults with cognitive disabilities, etc. I am somewhat ambivalent about this terminology because I view it as an underestimation of the agents in question, but I use it because I take my views to be in dialogue with Shoemaker's, and for ease of exposition.

<sup>2</sup> Smith, "On Being Responsible and Holding Responsible," 467-8.

<sup>3</sup> For a range of views on this topic, see Brink and Nelkin, "Fairness and the Architecture of Responsibility," Smith, "On Being Responsible and Holding Responsible," Wallace, *Responsibility and the Moral Sentiments*, Watson, "Responsibility and the Limits of Evil," Shoemaker, "Response-Dependent Responsibility" and McKenna, *Conversation and Responsibility*.

<sup>4</sup> Strawson, "Freedom and Resentment," 10.

<sup>5</sup> For example, Wallace's *Responsibility and the Moral Sentiments* and McKenna's *Conversation and Responsibility*. It is also worth noting here that though Strawson drew responsibility theorists' attention to the importance of the ways we treat those we regard as responsible agents, he never distinguishes between being responsible and holding responsible, nor uses the phrase 'holding responsible' at all.

<sup>6</sup> Some philosophers deny that the emotions play this central role in blame, and instead hold more cognitive views of blame (see, for example, George Sher's *In Praise of Blame*). These views are not my primary target in this paper. I am also not defending a particular view of blame. As such, I will remain agnostic as to whether holding someone responsible in the sense under discussion (that is, holding her to a norm) entails that she is the appropriate object of blame (in some sense or other) should she violate that norm. What I am claiming instead is that holding someone responsible need not entail the appropriateness of the negative reactive attitudes, even in response to norm violations.

<sup>7</sup> For more on this latter point, see section 3.

<sup>8</sup> Or, in the case of guilt, one does this with respect to oneself.

<sup>9</sup> "Enforcement" is sometimes associated with blame or punishment in response to norm violations. I am using this word more broadly, to mean roughly holding someone to a norm.

<sup>10</sup> Strawson, "Freedom and Resentment," 15.

<sup>11</sup> Strawson, "Freedom and Resentment," 15.

<sup>12</sup> We also characteristically take up the objective stance toward lower animals and inanimate objects, which is congruent with Strawson's view that we take up the objective stance when we can't sustain a certain kind of intimacy and reciprocity with that individual. Strawson claims that we can also take up the objective attitude toward anyone on a temporary basis for this kind of reason. As he writes, "We have this resource and can sometimes use it: as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity. Being human, we cannot, in the normal case, do this for long, or altogether" (Strawson 10).

<sup>13</sup> Strawson, "Freedom and Resentment," 10.

<sup>14</sup> In response to the incompatibilist determinist, Strawson insists that to take up the objective stance toward everyone all the time would be both inhuman and impossible. Because this argument is both well-known and not central to this paper, I will not rehearse it here.

<sup>15</sup> Strawson, "Freedom and Resentment," 9.

<sup>16</sup> Strawson, "Freedom and Resentment," 9.

<sup>17</sup> Watson, "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." 222-3.

<sup>18</sup> Wallace, *Responsibility and the Moral Sentiments*, 162.

<sup>19</sup> Note that Wallace’s conception of the reactive attitudes is much narrower than Strawson’s. Specifically, the reactive attitudes on his view are resentment, indignation, and guilt. I will discuss his argument for this claim below.

<sup>20</sup> Wallace, *Responsibility and the Moral Sentiments*, 27.

<sup>21</sup> Wallace, *Responsibility and the Moral Sentiments*, 21.

<sup>22</sup> Wallace, *Responsibility and the Moral Sentiments*, 23.

<sup>23</sup> Watson, “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme.” 222.

<sup>24</sup> I use the locution ‘affective blame’ following Hanna Pickard, but it is more or less interchangeable with blame understood in terms of the negative reactive attitudes, and contrasted with blame understood as merely a negative cognitive evaluation of the agent in question.

<sup>25</sup> Joffe and Levitt, *Conquering Depression: A Guide to Understanding Symptoms, Causes, and Treatment of Depressive Illness*, 39.

<sup>26</sup> Friedman, “When All Else Fails, Blaming the Patient Often Comes Next,” 1.

<sup>27</sup> Friedman, “When All Else Fails, Blaming the Patient Often Comes Next,” 1.

<sup>28</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1134-5.

<sup>29</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1134.

<sup>30</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1134.

<sup>31</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1135.

<sup>32</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1141.

<sup>33</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1150.

<sup>34</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1144-5.

<sup>35</sup> Smith, “Responsibility for Attitudes: Activity and Passivity in Mental Life,” 269.

<sup>36</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1149.

<sup>37</sup> Pickard, “Responsibility without Blame: Philosophical Reflections on Clinical Practice,” 1144.

<sup>38</sup> The point here is not that behaviors associated with disorders of agency are moral failings, but rather that the kind of accountability that is at issue in therapeutic relationships is garden-variety accountability, which is typically characterized as (one kind of) moral responsibility.

<sup>39</sup> This is not to say that the therapist only holds the patient responsible in a forward-looking sense. Assuming that the norm in question is a reasonable one and the patient is aware of it, there is no reason to think that she is not responsible for violating the norm in a backward-looking sense as well.

<sup>40</sup> Strawson, "Freedom and Resentment," 18.

<sup>41</sup> Strawson, "Freedom and Resentment," 10.

<sup>42</sup> Strawson, "Freedom and Resentment," 10.

<sup>43</sup> I use the language of treating as a person rather than the language of the participant stance because I believe the precise relationship between my view and Strawson's is complex. I share many of Strawson's methodological commitments, but I want to shift the focus of holding responsible away from the reactive attitudes.

<sup>44</sup> The point here is not that children or psychiatric patients are never the appropriate objects of blame. If we do in fact have normative expectations of marginal agents and those expectations are reasonable, it is doubtful that these agents are never to blame for living up to them. We may blame marginal agents less often because we are less sure that our expectations are reasonable, and the tone of our blame typically is (and should be) softened by recognizing their vulnerability. But the same can be said of the grieving employee considered below, who is not a marginal agent.

<sup>45</sup> This example has an affinity with one Angela Smith discusses in "On Being Responsible and Holding Responsible." Smith writes, "If a good friend of mine is under a lot of stress, for example, I may not 'hold' her responsible, in the sense of actively blaming her, for some insensitive comments she makes to me. I can judge both that she is responsible for her comments, and that she is open to legitimate moral criticism for them (because they are hurtful). But given the circumstances, I may decide that it would be uncharitable for me to take up attitudes of anger and resentment, or to explicitly reproach her in any way. In making such a judgment, however, and in renouncing these attitudes and responses, I need not think that my friend is not really responsible, or not really at fault, for her behavior." (Smith (2007) 470). According to Smith, therefore, judging someone to be responsible (and perhaps even culpable) need not entail the justifiability of the negative reactive attitudes. In this I agree with Smith, but on my view holding someone responsible is fundamentally about holding her accountable with respect to a norm, not merely judging her to be responsible or attributing an action "...to her in a way that makes moral appraisal, in principle, appropriate." (Smith (2007) 471). I also disagree with Smith that holding someone morally responsible in a backward-looking sense is generally synonymous with blaming her (Smith (2007) 468).

<sup>46</sup> More precisely, praising or crediting someone with respect to some norm counts as holding her responsible if, had she violated that norm, she would also have been held responsible (e.g., by being blamed or punished in some way).

<sup>47</sup> Gary Watson, for example, famously argues that responsibility as accountability is not the only sense of responsibility relevant to our responsibility practices; there is also responsibility as attributability, which he describes as follows: "conduct can be attributable or imputable to an individual as its agent and is open to appraisal that is therefore appraisal of the individual as an adopter of ends. Attributability in this sense *is* a kind of responsibility. In virtue of the capacities in question, the individual is an agent in a strong sense, an author of her conduct, and is in an important sense answerable for what she does" (Watson (1996) 229). My interest here is not in denying that attributability could also be a form of responsibility, but in developing a

view of responsibility as accountability that captures the full array of cases in which we in fact hold each other accountable.

## REFERENCES

- Brink, David O., and Dana K. Nelkin. "Fairness and the Architecture of Responsibility." *Oxford Studies in Agency and Responsibility*, no. 1 (2013): 284–313.
- Friedman, Richard A. "When All Else Fails, Blaming the Patient Often Comes Next." *The New York Times*, October 21, 2008. [www.nytimes.com/2008/10/21/health/21mind.html](http://www.nytimes.com/2008/10/21/health/21mind.html).
- Joffe, Russell T. and Anthony J. Levitt. *Conquering Depression: A Guide to Understanding Symptoms, Causes, and Treatment of Depressive Illness*. BC Decker Inc., 2000.
- McKenna, Michael. *Conversation and Responsibility*. Oxford University Press, 2012.
- Pickard, Hanna with commentary by Lisa Ward. "Responsibility without Blame: Philosophical Reflections on Clinical Practice." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by KWM Fulford, Martin Davies, Richard Gipps, George Graham, John Sadler, Giovanni Stanghellini, and Tim Thornton, 1134–54. Oxford University Press, 2013.
- Sher, George. *In Praise of Blame*. Oxford University Press, 2005.
- Shoemaker, David. "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame." *Philosophical Review* 126, no. 4 (2017): 481–527.
- . *Responsibility from the Margins*. Oxford University Press, 2015.
- Smith, Angela. "On Being Responsible and Holding Responsible." *Journal of Ethics* 11, no. 4 (2007): 465–84.
- . "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236–71.
- Strawson, P.F. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. Routledge, 2008.
- Wallace, R. J. *Responsibility and the Moral Sentiments*. Oxford University Press, 1994.
- Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Agency and Answerability: Selected Essays*, 219–59. Oxford University Press, 2004.
- . "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1995): 227–48.