

Georgia State University

ScholarWorks @ Georgia State University

Philosophy Theses

Department of Philosophy

8-8-2017

Too Tired to be Fair: Reactive Attitudes and Irrelevant Influences

Amanda Haskell

Follow this and additional works at: https://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

Haskell, Amanda, "Too Tired to be Fair: Reactive Attitudes and Irrelevant Influences." Thesis, Georgia State University, 2017.

doi: <https://doi.org/10.57709/10423276>

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

TOO TIRED TO BE FAIR: REACTIVE ATTITUDES AND IRRELEVANT INFLUENCES

by

AMANDA HASKELL

Under the Direction of Eddy Nahmias, PhD

ABSTRACT

Reactive attitudes are distinctively moral emotions that occur when a moral harm has occurred. Recent studies in moral psychology suggest that our reactive attitudes may be influenced by factors extraneous to moral evaluation, such as hunger, sleep deprivation, and negative moods. I argue that these influences lead us to sanction unfairly. Even though reactive attitudes may be a natural response to perceived moral wrongdoing, we cannot justifiably inflict undeserved harm. However, if we can learn to recognize and eliminate the effects of these irrelevant influences, then we can use our reactive attitudes productively in holding others morally accountable.

INDEX WORDS: Reactive attitudes, Emotions, Accountability, Responsibility, Moral psychology, Ego depletion

TOO TIRED TO BE FAIR: REACTIVE ATTITUDES AND IRRELEVANT INFLUENCES

by

AMANDA HASKELL

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2017

Copyright by
Amanda Marie Haskell
2017

TOO TIRED TO BE FAIR: REACTIVE ATTITUDES AND IRRELEVANT INFLUENCES

by

AMANDA HASKELL

Committee Chair: Eddy Nahmias

Committee: Andrea Scarantino

Eyal Aharoni

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

DEDICATION

To my parents, who have been incredibly supportive and patient for all the time I spent working on this thesis instead of keeping in touch with them.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Eddy Nahmias, for consistently providing the feedback and guidance that fostered my intellectual growth and made this project possible; Andrea Scarantino for pushing me to write in a more focused and organized manner; Eyal Aharoni for giving me valuable experience doing psychological research and helping me learn how to handle and analyze data; Michael De Vivo for his continual support and feedback; and the rest of my friends for making this whole experience more rewarding and enjoyable.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
1 INTRODUCTION.....	1
2 REFINING STRAWSON’S ACCOUNT OF THE REACTIVE ATTITUDES....	3
2.1 The evaluative component.....	7
2.2 The bodily component.....	8
2.3 The motivational component.....	10
3 RESPONSIBILITY AND ACCOUNTABILITY.....	12
4 INCIDENTAL INTERNAL STATES AND REACTIVE UNRELIABILITY ...	21
4.1 Hunger.....	23
4.2 Sleep deprivation.....	24
4.3 Negative mood	27
4.4 Objections	30
5 CONTROLLING OUR REACTIVE ATTITUDES.....	35
6 CONCLUSION	43
REFERENCES.....	46

1 INTRODUCTION

Imagine coming home after a tiring day of work to find that some of your spare cash is missing. You assume your roommate stole your money and feel resentment towards him. You immediately go to his room and lash out at him for betraying your trust. Instead of thinking to ask for an explanation, you angrily kick him out of the apartment. Even if you can't relate to this particular case, it is unlikely that you have never reacted emotionally to another's moral transgression. And if someone has wronged you on the wrong day—after missing a meal or a good night's sleep—you may have noticed that your reactive emotions felt stronger and more difficult to control. In this paper, I seek to explore the variability of our moral reactive attitudes and the implications of that variability for our sanctioning responses.

When P.F. Strawson argued for the importance of our reactive attitudes in holding others morally responsible, he rightly acknowledged that these attitudes should occasionally be regarded with mistrust. However, in consideration of recent research, the reliability of our reactive attitudes may be undermined by factors more mundane than Strawson himself considered, i.e., “self-deception...guilt transference, unconscious sadism and the rest” (Strawson 1962: 203). Namely, recent empirical evidence suggests that our reactions to moral transgressions are impacted by various physical states—such as hunger and sleep deprivation—and incidental bad moods.

Although Strawson only briefly mentions psychological studies that reveal the dark side of the reactive attitudes, recent research forces us to question the reliability of the reactive attitudes more than ever. I argue that we cannot take our reactive attitudes for granted in cases where they are prone to regularly lead us to inflict undue harm on others, a concern that Strawson and others have failed to explore.

In Section 2, I elaborate on Strawson's account of the reactive attitudes, focusing on the role that reactive attitudes play in our understanding of moral responsibility. I then situate the reactive attitudes within a theoretical framework of emotion in order to paint a clearer picture of how they affect our treatment of others. In Section 2, using a distinction made by Gary Watson, I clarify the notion of *moral responsibility* relevant to my discussion of the reactive attitudes. I will focus on responsibility as accountability—which justifies the imposition of moral sanctions (such as punishment)—and show precisely how reactive attitudes may lead us to sanction unjustly. In Section 4, I present empirical evidence that suggests that our moral reactive emotions vary in strength due to common internal states, such as incidental sadness, sleep deprivation, and hunger. This variation may lead us to sanction when an agent does not deserve it or sanction more severely than the agent deserves.

I conclude, when the underlying motives of our sanctioning behaviors go unchecked, we cannot rely on our reactive attitudes to motivate fair sanctions. However, I do not take this conclusion to mean that our reactive emotions are always unviable for holding others morally accountable. I will argue that if individuals and institutions can learn to assess reactive emotions before they affect our sanctions toward others—to make sure they are not providing bad information due to irrelevant influences—then they may be helpful in later assessing the agent's action and implementing, if warranted, a proper sanctioning response. If we use various methods of self-control, I believe it is possible to use reactive attitudes productively in our sanctioning decisions.

2 REFINING STRAWSON'S ACCOUNT OF THE REACTIVE ATTITUDES

In “Freedom and Resentment” (1962), P.F. Strawson introduced the reactive attitudes in part to argue that the truth or falsity of determinism is irrelevant to the way we hold each other morally responsible.¹ Although the ‘pessimists’—those who endorse a merit-based notion of moral praise and blame—may believe that determinism invalidates these practices, Strawson argues that the moral reactive attitudes cannot be eradicated even if one believes that people do not have free will.² In other words, even if someone believed that no one has the kind of free will necessary to justify merit-based punishment, that belief would not prevent her from reacting in ways indicative of moral praise or blame.

Strawson argued that experiencing a reactive attitude *just is* holding someone morally responsible. When we acknowledge someone as a fellow member of the moral community, we implicitly make certain demands of her: demands for good will or, at least, for the absence of ill will. Strawson claims that “the making of the demand is the proneness to such attitudes”, such that we do not experience reactive attitudes in the absence of these sorts of demands (1962: 200). For example, we demand that others respect our belongings by not damaging or stealing them. When you express resentment to your thieving roommate, it is implied that you believe him to be responsible for violating that demand. Strawson notes that “our personal feelings and reactions depend upon, or involve our beliefs” about the agent’s actions and intentions towards us (1962:186). If it is clear that the upsetting action was performed accidentally—like if your roommate had merely moved your money while cleaning—then you may react angrily but will not blame him because no demand was violated.

¹ Strawson identifies resentment, indignation, disapprobation, guilt, forgiveness, hurt feelings, and some types of love as reactive attitudes (1962).

² As I will elaborate on later, a merit-based account of moral responsibility is backward-looking and focuses on what an agent deserves for the action they have performed. It is not primarily intended to change their future behavior.

Strawson's contribution to the debate on determinism and moral responsibility is largely practical: reactive attitudes are an integral feature of our social relationships and are too deeply engrained and valued to be eliminated. We care about the actions of others—especially those close to us—insofar as they reflect good or ill will toward us. If we could eliminate our reactive attitudes and enter into a detached state towards everyone, we would also eliminate many of the feelings and behaviors that give value to our personal relationships. For instance, we could not react with gratitude toward those who help us, resent those who do us harm, or react indignantly towards social injustices. Strawson notes that “in the absence of any forms of these attitudes it is doubtful whether we should have anything that we could find intelligible as a system of human relationships, as human society” (1962: 202).

Although Strawson provides good reason to accept that the reactive attitudes are a critical element of human socialization, he does express concern over psychological studies that reveal the dark side of our reactive attitudes. While he does not believe the reactive attitudes should be (or could be) rejected entirely, he accepts “the possibility and desirability of redirection and modification of our human attitudes in light of these studies” (Strawson 1962: 203). However, he fails to explore this consideration, adding it only as a qualification at the end.

In the time since Strawson originated this theory, a large body of scientific literature has suggested that the reactive attitudes are susceptible to numerous irrelevant influences. Due to these studies, I wish to further pursue Strawson's consideration that our reactive attitudes should be modified to lessen the effect of negative psychological influences. I agree with Strawson that the reactive attitudes are a central feature of our social behavior. However, contra Strawson, I argue that the empirical evidence put forth warrants a thorough reconsideration of the role of the reactive attitudes in our sanctioning practices. If our reactive attitudes lead us to treat others

unfairly, then we have good reason to significantly revise our behavior rather than taking our reactive attitudes at face value.

In order to evaluate the psychological literature and its implications for Strawson's reactive attitudes, a more comprehensive analysis of the reactive attitudes is needed. Strawson himself notes that it would be appropriate to refer to the reactive attitudes as *moral sentiments*, which indicates that he views them as akin to emotions rather than mere judgments (Strawson 1962: 202). I believe that analyzing the emotional nature of the reactive attitudes is critical to understanding how these attitudes may be subject to irrelevant influences, how these influences detrimentally impact our behavior and social relationships, and how we can modify our reactive attitudes to avoid the negative effects of these influences.

For the sake of this argument, I will focus on the negative reactive attitudes exclusively because of their connection with negative treatment and sanctions. I will focus my analysis on an attitude identified as central to moral responsibility by Strawson: resentment. However, this account will transfer smoothly to most of the attitudes named by Strawson, but some cases are unclear. For example, forgiveness can be regarded as a process or action rather than an emotion. It is unclear whether forgiveness involves a distinctive bodily state, such as the feeling of anger or sadness. However, this is not a major problem: the negative reactive attitudes are of greater concern due to their relationship with blame and punishment.

To evaluate the reactive attitudes as emotions, I will draw on the Attitudinal Account of emotion proposed by Julien Deonna and Fabrice Teroni (2012). The Attitudinal Account allows us to fill in a gap left by Strawson by providing a clear picture of how the reactive attitudes function as emotions. Although the literature is far from settling on a definitive account of emotion, the Attitudinal Account is a recent theory that puts emphasis on how emotions motivate

actions. By investigating the connection between emotion and action, we can better understand how our reactive attitudes lead us to act in response to perceived moral violations. The Attitudinal Account's emphasis on the evaluative properties of objects allows us to get a clear picture of what it means for a reactive attitude to be fitting or unfitting and, by extension, what it means for a sanction motivated by a reactive attitude to be fair or unfair.

Notably, this analysis deviates from Strawson's account by maintaining that belief is not constitutive of reactive attitudes. Where Strawson maintains that our reactive attitudes depend on or involve beliefs, the Attitudinal Account does not consider beliefs or other cognitive states to be a precursor or integral element of emotions. Rather, on this account, it is most accurate to say that having a reactive attitude leads to a belief about an agent's blameworthiness for violating a demand. The reactive attitude itself is best seen as an embodied response to an act of ill will. To distinguish my treatment of the reactive attitudes from Strawson's, I propose to understand the belief-like component of reactive attitudes as *subdoxastic states* (Stich 1978). Such states play a proximate role in the production of genuine beliefs but lack characteristic properties of belief, most importantly grounding inferences.

I will maintain the language of *fittingness* for the duration of this paper, in order to avoid claiming that emotions can be correct or incorrect. Where a belief may be true or false, an emotion may be fitting or unfitting depending on whether it accurately corresponds to the evaluative properties of its object (D'Arms & Jacobson 2000). Using the Attitudinal Account as a foundation, I will break down the reactive attitude of resentment into three components: evaluative, bodily, and motivational. Deonna and Teroni do not use these distinctions specifically, but these elements of emotions nonetheless figure centrally into their account.

2.1 The evaluative component

On the Attitudinal Account, emotions are “distinctive evaluative attitudes,” which is consistent with the claim that the reactive attitudes occur only when we have perceived that an agent has violated a demand (Teroni & Deonna 2012: 78). We judge the fittingness of an emotion by looking to the “corresponding evaluative property” in the object of the emotion (Teroni & Deonna 2012: 81).³ On the Attitudinal Account, the reactive attitudes should accurately track the relevant evaluative properties of an object. Consider fear, the emotion that tracks the dangerousness of an object. Let us suppose that the dangerousness of an object can be rated from 1-10. Because sharks score a solid 8 on the dangerousness scale, a high-level fear response is fitting. Because houseflies do not even deserve a 1 on the dangerousness scale, any fear response would be unfitting.

To get an idea of what it would look like to have a fear response that *does not* fittingly track relevant evaluative properties, imagine staying up all night to watch *The Exorcist*. Still feeling unsettled and somewhat disoriented from sleep deprivation, you start your day by sipping coffee outside on the porch. Suddenly, your neighbor’s dog barks and starts running towards you, causing you to drop your coffee and jump inside. Your many past experiences with this dog have indicated that it poses no danger to you whatsoever. However, due to the movie you watched and lack of sleep, you have an extremely fearful reaction to the dog. This reaction does not track the relevant evaluative properties of the dog (i.e., the dog’s danger to you) making it an unfitting emotional reaction. Although you might have been surprised, such a strong fear reaction would not have occurred in the absence of the negative mood imparted by the movie and the effects of sleep deprivation.

³ Although Deonna and Teroni use correctness, I do not believe my substitution of fittingness will problematize my use of their Attitudinal Account

In the case of resentment, we must judge its fittingness by evaluating whether an agent's action violated a moral demand in a way indicative of ill will. If someone forcefully shoves you out of hatred, your resentment tracks their ill will: their disrespect toward you is a relevant evaluative property and your resentment is a fitting response to it. However, if someone accidentally shoves you because they slipped on a wet floor, their action did not stem from any ill will towards you. In light of this, a resentful response would be inappropriate because it would not correspond to any relevant evaluative property in the clumsy person. Any experience of resentment that does not correspond to actual ill will in the elicitor is unfitting. It may be the case that you immediately feel resentful toward the person who shoved you if you do not realize their movement was accidental. However, this response is only problematic if it does not dissolve when the individual explains that their action was an accident; your resentment should lose its apparent fittingness in light of their explanation and the emotion should subside.

2.2 The bodily component

According to the Attitudinal Account, distinct evaluative attitudes are identical with our body's felt inclination to act in a certain way in response to a certain stimulus. Thus, in the first instance, an emotion's fittingness to a particular evaluative property E (e.g., danger) is grounded the body's selective causal sensitivity to E (e.g., heightened stress response) over other evaluative properties. The experience of a reactive attitude, on the Attitudinal Account, involves "awareness of one's body adopting a specific stance towards an object or being poised to act in given ways in relation to an object" (Teroni & Deonna 2012: 79). Resentment may lead to physiological changes such as increased heart rate, perspiration, and the release of cortisol or adrenaline (neurotransmitters associated with stress responses). Even if you do not always notice your pulse quicken when you are angry, you surely know what it *feels like* to be angry. This

component is particularly relevant for resentment, as many are likely to have a strong bodily response to personal injury. The same can be said for guilt, which is frequently described as persistent and sickening until the guilty agent makes up for their wrongdoing. Another attitude included by Strawson—indignation—may not have a strong bodily component. This does not mean we have no emotional response when we witness injustice; this response just has a higher likelihood of being affectively cool, as we are not direct participants in the offensive situation.⁴

Notably, the experience of an emotion may affect how we perceive and think about the elicitor of the emotion. If you are angry with your spouse, other agitating behaviors of hers—such as not cleaning up after herself or forgetting to fill the car with gas—may become salient when you would have overlooked them normally. The feeling of resentment may lead us to focus on the negative characteristics of the elicitor and withdraw our favorable opinion of them, at least until we are offered an exonerating explanation, an apology, or amends.

Importantly, we should note that the bodily experience of an emotion does need not be *expressed*. Most know the experience of suppressing tears when sad or holding back from shouting a nasty insult during a heated argument. Although emotions poise us to act in a certain way and often create a desire or impulse to act in that way, most of us are able to resist these reactive behaviors and can improve our ability to do so. If this were not the case, anger management therapy could not aid people in controlling their temper—but evidence strongly suggests such methods can be effective (Beck & Fernandez 1998).

⁴ Relatedly, Deonna and Teroni allow for emotions for which no informative embodied characterization is forthcoming, writing: “Those skeptical of the possibility of coming up with such informative characterizations for some emotions – regret? pride? – might still embrace [the Attitudinal Theory] and try to account for these allegedly recalcitrant emotions in terms of felt, yet non-embodied, attitudes” (p. 81).

2.3 The motivational component

The bodily component of the reactive attitudes also motivates us to act in certain ways towards the individual who evoked the attitude. Feeling fear does not merely indicate that we have perceived that a stimulus is dangerous. Further, fear prepares the body to deal with the dangerous stimulus (via fighting or fleeing) and puts us in a mindset conducive to pursuing these options. When the fight-or-flight response is triggered, we feel strongly motivated to undertake one of those actions. If acknowledging danger did not involve this bodily component, we would be less prepared to deal with the stimulus physically and less inclined mentally to pursue an appropriate course of action. The same can be said for all reactive attitudes. The fact that indignation has a weaker bodily component may partially explain why we act on it less frequently. It is easier to overlook the maltreatment of another than the maltreatment of oneself.

Although the bodily experience of an emotion need not lead to significant outward behavioral expressions, it frequently does. Unlike the harmful behaviors that result from uncontrolled anger, we tend to believe that many bodily motivated behaviors are appropriate responses to the elicitor. Most clearly, running away in fear is a beneficial response to a dangerous stimulus. Similarly, most find it justified to treat someone differently in response to personal injury. Strawson emphasizes that the negative reactive attitudes “tend to inhibit or at least to limit our goodwill towards the object of these attitudes” (2001: 200). For instance, because we have withdrawn our favorable opinion of the elicitor of resentment, we may lose our motivation to interact with them congenially and may wish to avoid them altogether. These sorts of actions demonstrate a retraction of goodwill. If a friend or partner has ever subjected you to the silent treatment after a fight, you have experienced the effects of diminished good will.

However, our reactive attitudes may motivate us to do more than merely withdraw our good will. If we perceive that another has acted out of ill will towards us, we might feel inclined to express ill will in response. We may alter our body language and facial expressions to convey a negative emotional response, though these changes can occur unconsciously when resentment is experienced. We may also speak in a harsh tone, yell, use insulting language, or act belligerently toward the individual themselves or their belongings. Further, our reactive attitudes can lead us to make hasty sanctioning decisions when more deliberation would be appropriate. For example, an angry parent may prohibit their child from attending an event before hearing the child's explanation of his actions. Such a sanction may have been completely undeserved, as the child's explanation may have rendered the reactive attitude unfitting.

In sum, reactive attitudes can motivate our behavior toward others in myriad ways. Although some of these behaviors may be consciously chosen as a sanctioning response, they often occur more immediately, before any such decision has been made. This means that unfitting reactive attitudes—those that do not accurately track the relevant evaluative properties of the object—can strongly affect our sanctioning responses. For one, we may be so overcome by the reactive attitude that we do not allow the agent a chance to explain their actions before we subject them to sanctioning behavior. Additionally, we may impose more severe sanctions than the individual deserves.

These distinctions are crucial for judging when a reactive attitude is fitting. The evaluative component supplies the basic criteria for assessing when a reactive attitude is hitting its target: when it accurately corresponds with the quality of an agent's will. The bodily component provides insight into how reactive attitudes alter our perception of an agent and motivate various motivational responses. The motivational component helps to demonstrate how

the reactive attitudes can lead us to cause undue harm to another. By gaining a more comprehensive understanding of the reactive attitudes as multifaceted emotions, we can more readily identify remedies for the unfairness introduced by irrelevant influences and justify the role of the reactive attitudes in holding others morally responsible.

In the following section, I will address issues regarding the fittingness of our reactive attitudes and the fairness of our sanctioning responses. I will first provide a discussion of moral responsibility and justificatory criteria for imposing sanctions to gain a better understanding of what makes a sanction unfair. I will then investigate the ways in which our unfitting emotions can prevent us from meeting the criteria specified for fair sanctioning.

3 RESPONSIBILITY AND ACCOUNTABILITY

Strawson's account is a merit-based account of responsibility in which the experience of a reactive attitude itself is constitutive of holding someone responsible. In experiencing a reactive attitude, we are recognizing the agent's moral wrongdoing (evaluatively), experiencing their blameworthiness (bodily), and responding to their ill will (motivationally). As outlined earlier, in the case of resentment, the evaluative dimension consists of the evaluation that an individual has violated a moral demand in a way indicative of ill will. This evaluation is constitutive of the bodily experience of their perceived wrongdoing—the feeling that the agent has wronged you. In turn, this bodily experience of their blameworthiness motivates the imposition of moral sanctions: one feels that the agent *deserves* sanctions because of their actions (Fehr & Gächter 2011).

To reinforce this account, consider the population that we take the objective stance towards, according to Strawson. He insists that we can suspend our reactive attitudes (i.e. take

the objective stance) when the perpetrator is psychologically abnormal; we see them “as an object of social policy, as a subject for what, in a wide range of senses, might be called treatment” (Strawson 2001: 190).⁵ This perspective corresponds to a consequentialist account of responsibility, as sanctions serve only deterrent or curative functions and treatment is not a thing that can be deserved (Lewis 1953: 224). We do not think these individuals *deserve* sanctions because they cannot properly function as members of the moral community; they simply do not have the sort of moral understanding or control over their behavior necessary to be the fitting targets of reactive attitudes. The experience of a reactive attitude includes the evaluation that an agent *did* have sufficient understanding of her moral obligations and sufficient control over her decision and, because of that, *deserves* to be treated a certain way. This treatment is backward-looking and, unlike consequentialist accounts, not primarily intended to have any particular effects on the agent’s future behavior.

Now, one may make the point that people regularly *do* experience reactive attitudes towards psychologically abnormal individuals who commit certain crimes. If we experience reactive attitudes towards psychologically abnormal individuals who cannot be properly held responsible, then our reactive attitudes generally do not contain the evaluation that the agent is morally responsible. I will suggest a few responses to show that the reactive attitudes only occur if we perceive the agent to be properly responsible.

First, if the reactive attitudes are emotional responses that are triggered by the violation of a moral demand, then they may still initially be triggered when an individual knows the perpetrator is not a proper member of the moral community. Imagine that you are submerged in the ocean in a completely foolproof shark tank for the first time. Although you know that the

⁵ He is primarily referring to those with certain types of mental illness, such as psychopathy or schizophrenia, where the agent is unable to grasp morality or may be fully delusional (Strawson 1962).

sharks cannot hurt you, you will likely still feel fear when a Great White swims toward you. The case is similar with the experience of resentment towards a mentally ill person—like a delusional schizophrenic—who has acted immorally. We can still assume the objective stance *after* this initial emotional experience because we know it is not fitting for the case at hand. Even if the bodily experience of resentment is not fully extinguished, we can prevent ourselves from expressing this resentment through action.

Alternatively, some who experience reactive attitudes in this context may incorrectly view the perpetrator as a member of the moral community due to a failure to genuinely comprehend their psychology. The more an individual can understand the impulse to perform a particular immoral action, the more they may project their own psychology onto someone who actually performs that action. It then becomes easy to blame another agent for performing an action that they themselves have successfully resisted. For instance, someone who occasionally fights against the desire to shoplift is more likely to experience a reactive attitude toward a kleptomaniac—they cannot bring themselves to believe that the kleptomaniac’s mind is significantly different from their own. It may be easier to assume the objective stance toward a serial killer, as they are so far removed from the typical human experience that this projection does not factor in. However, cases involving horrific acts may still evoke empathy for the victims and moral disgust for the act itself.

Now, to gain a greater understanding of how experiencing a reactive attitude is constitutive of holding an agent responsible, we must clarify what is meant by *holding responsible*. To do this, I will focus on a distinction made by Gary Watson (1996): responsibility as attributability (hereafter “holding responsible”) versus responsibility as accountability

(hereafter “holding accountable”).⁶ In holding an agent responsible, we attribute bad conduct to an agent: we are justified in making them explain their behavior because they were the author of the behavior. However, we are not justified in imposing moral sanctions because we have not yet ascertained whether the agent is genuinely blameworthy.

Watson’s notion of responsibility as *accountability* goes beyond mere attributability: when we hold someone accountable for their actions, we are imposing a liability to sanction them. Because sanctions (e.g. punishment) impose adverse treatment (including potential suffering) on the agent, holding someone accountable requires significantly higher justification than merely having them answer for their conduct. Watson specifies two justificatory criteria: we must be entitled to make demands upon others (as liability to sanctions follows directly from the making of a demand) and the agent must have a reasonable opportunity to avoid said sanctions (Watson 1996: 237).

One way that we can fulfill the latter criterion is by allowing an agent to explain her conduct, as it may be that the agent made a genuine mistake or was coerced to perform the action. In those cases, failure to demand an explanation might lead one to impose a wholly undeserved sanction. Even in cases where the agent did act out of ill will, an explanation will provide information to factor into the sanctioning decision. Watson concedes that the first criterion is more problematic, as it is difficult to evaluate what gives someone the authority to impose sanctions on another. Given Watson’s criteria, I maintain that, minimally, an agent must be able to make reasonable demands and assure their sanctioning response fits the violation of those demands.

⁶ Watson does not use this terminology, but I believe it provides a more precise framework to evaluate the role of reactive attitudes in our practices of moral responsibility.

It is crucial to clarify what counts as a sanction. Watson uses the term ‘sanction’ to refer to “certain adverse or unwelcome treatment” that an agent is liable to when he violates a moral demand (Watson 1996: 236-237). Watson accepts that “being subject to blaming attitudes” counts as a sanction in itself—even when the attitudes are not explicitly expressed—because they involve dispositions to treat others in generally unwelcome (or less friendly) ways (Watson 1996: 238). As elaborated above, the experience of a reactive attitude changes how we perceive the agent and disposes us to treat them differently. Even if we do not straightforwardly express the emotion—by displaying a facial expression, acting aggressively, or choosing a deliberate sanction—we may still withdraw good will, treat the agent less cordially, and see them in a worse light. According to Watson, these changes count as subjecting the agent to adverse treatment.

Watson also points out that we can hold the dead accountable and blame them for faulty actions, but cannot actually sanction them. In order to maintain that the dead can be unfairly blamed, Watson argues that “one’s blaming attitudes are unfair if it would be unfair for whatever reason to subject others to the adverse treatment to which one’s attitudes dispose one” (Watson 1996: 239). This move allows us to claim that we can unfairly blame people even when treating them adversely is impossible. When we sanction an agent, we are doing so unfairly if they do not deserve the sanction at all—because they weren’t truly accountable for the action in question—or if the sanction imposed is more severe than they deserve. From this, it follows that our reactive attitudes are unfair when they dispose us to subject agents to adverse treatment that they do not deserve. And our reactive attitudes dispose us to subject agents to undeserved treatment when they are unfitting.

An example can clarify these concepts further. Recall the example of your having just returned home to find that your roommate stole your cash. You feel resentment in response to their action and ask him for an explanation. Let's say that your roommate now reminds you that, last weekend, you gave him permission to borrow some of the money you keep in the house. You recall this forgotten interaction and your feeling of resentment subsides immediately because he expressed no ill will by taking the money. Alternatively, suppose your roommate explained that he didn't care about respecting your belongings and would never have mentioned it if you hadn't noticed. Now you seriously consider evicting him from your apartment, as he may be properly held accountable for their conduct, as it indicates disrespect and a complete lack of good will: your roommate has violated your reasonable demand that he respects your property.

In this example, Watson's second criterion is certainly met: in asking your roommate to explain his conduct, you gave him a reasonable chance to avoid sanctions. In the instance where you actually gave him permission to borrow your money, his explanation rightfully exempted him from being held accountable. I believe it is clear that Watson's first criterion was met as well. Stealing is readily acknowledged as an immoral action and we reasonably expect others to respect our belongings. Additionally, as the owner of your apartment, you know that you have the right to evict your roommate after they commit a serious violation of the lease.

Let us consider one more variation of the scenario. Suppose that you come home exhausted and grumpy over your day at work. When you see that your money is gone, you are so overcome with resentment that you fail to ask for an explanation. Unbeknownst to you, if you had requested an explanation, your roommate would have told you that you had given him permission to borrow the money. Instead, you yell at him for disrespecting your property and

threaten to kick them out of the apartment. Your roommate is clearly upset by your reaction and tries to explain as they fight back tears. Once you remember having given him permission, you feel guilty for having threatened them so harshly.

As evidenced in the example above, our unfitting reactive attitudes can prevent us from meeting the criteria necessary to justify sanctioning responses. Although your roommate was answerable for taking the money, he was not accountable for stealing because he had permission to take the money. The strength of your reactive attitude prevented you from uncovering a relevant explanation and you subjected your roommate to a sanctioning response he did not deserve. Even if you had not actively expressed your resentment, you may have felt disposed to ignore him and refrain from treating him kindly, putting strain on your living situation. In that case, your unexpressed reactive attitude was still an unfair sanctioning response.

We can also imagine cases where an agent should be held accountable but the strength of the reactive attitude is unfitting and leads you to sanction too severely. Perhaps upon finding that your roommate (actually) stole your money, you punch him unconscious. Even though his conduct was genuinely faulty, it would be difficult to justify such a violent sanction. Strawson himself emphasizes that the reactive attitudes should correspond with the magnitude of the ill will expressed by the agent:

[R]esentment tend[s] to inhibit or at least to limit our goodwill towards the object of [this] attitude, tend[s] to promote an at least partial and temporary withdrawal of goodwill; [it] does so in proportion as [it] is strong; and [its] strength is in general proportioned to what is felt to be the magnitude of the injury and to the degree to which the agent's will is identified with, or indifferent to, it. (Strawson 1962: 200)

In the following section, I will outline studies that suggest that our reactive attitudes do not fittingly respond to the relevant features of others' moral actions: they are influenced by irrelevant factors such as hunger and incidental mood. Before presenting the evidence, let me clarify here why these factors are irrelevant. On Watson's account, the making of a moral demand necessarily entails a liability to sanctions: "To require or demand certain behavior of an agent is to lay it down that unless the agent so behaves she will be liable to certain adverse or unwelcome treatment" (Watson 236-237). On Strawson's account, the reactive attitudes are constitutive of holding an agent accountable because they only occur in response to perceived violations of moral demands indicative of ill will; if we did not perceive the agent to be accountable for the violation, we would not experience a reactive attitude. If an emotion was not triggered by the perceived violation of a moral demand, Strawson would not consider it a reactive attitude. Whenever we experience a reactive attitude towards someone, we are holding them morally accountable. To respond to the violation appropriately, we must guarantee that our reactive attitudes are fitting, in that they accurately track the ill will expressed in the violation. If our reactive attitudes are not fitting, then we cannot fairly hold someone accountable for their action because we do not have an accurate assessment of that action.

The effects of factors like hunger and sleep deprivation are irrelevant to moral evaluation because they are wholly unrelated to the quality of an agent's will. Ideally, at what I will suggest we consider our "baseline" state—when we are well-rested, well-fed, and not in a bad mood—our reactive attitudes can accurately track the ill will expressed in an action. At baseline, we experience a certain level of emotional and physiological arousal and can control our behavioral impulses accordingly until we can better assess the situation by gaining additional information. Basically, baseline is a state in which our executive capacities are not significantly affected by

irrelevant influences, much like sobriety refers to a state in which we are not significantly affected by a psychoactive substance.

I am assuming that the baseline state generally produces emotional reactions that properly track relevant evaluative properties and are roughly similar across a population. I believe this baseline state can be compared to other biological features of humans, such as eyesight. Our baseline for proper vision is 20/20 and we seek to correct eyesight that falls below baseline because it indicates worsened visual performance. Additionally, there are some things that temporarily disrupt the baseline state of our eyesight, such as sleepiness or alcohol causing blurry vision. In these cases, we have no problem saying that our eyesight is impaired until we return to the baseline state. Similarly, we have a physical baseline state in regards to our emotional reactions and deviations that bring us below this baseline state—influences that impair relevant psychological capacities—impair the accuracy of our emotional reactions.

However, our moral reactive attitudes are importantly different from something like eyesight in that different people have different moral motivations. That is to say, each individual weighs moral values differently and has a unique profile of both backward and forward-looking concerns when considering sanctions. Although Strawson's account is backward-looking, I do not believe that precludes my discussion from acknowledging forward-looking reasoning. Given the high degree of variability in individual moral reasoning, it is difficult to say what the baseline state should look like. However, we can target clear cases of moral reasoning going awry, as in the cases where irrelevant influences alter an individual's moral emotional reactions and reasoning.

Irrelevant influences can alter the strength of our reactive attitudes by heightening the level of arousal produced by the reactive attitude and impairing our ability to control and

evaluate our response. In these cases, the reactive attitude is overwhelming and prevents us from better evaluating the situation—we may not seek out additional evidence and may not be able to integrate all of the evidence into an accurate evaluation. Although our being sleepy has nothing to do with assessing the quality of an agent’s will, it influences our emotional reaction to the agent, such that our reaction differs from how we would respond at baseline.

Merit-based accounts of responsibility are defined by giving an individual their just deserts. The variability in our reactive attitudes prevents us from guaranteeing that our sanctioning responses are accurately informed. Due to their unreliability, the reactive attitudes may lead us to impose unfair sanctions. Because of this, I argue that we must take steps to assure the fittingness of our reactive attitudes before imposing sanctions on an individual.

4 INCIDENTAL INTERNAL STATES AND REACTIVE UNRELIABILITY

A number of studies suggest that incidental emotions and physical states lead people to react more strongly and sanction more severely than they would at baseline. I will focus on three different states: hunger, sleep deprivation, and incidental negative moods. I suggest that these states all share a common thread: ego depletion. Although Roy Baumeister’s (2007) original model of ego depletion portrays self-control as a limited physical resource, I will frame my discussion using Michael Inzlicht & Brandon Schmeichel’s (2012) updated model that focuses on alterations in our attentional and motivational capacities.⁷ According to the motivational model, self-control is not a definitively limited physical resource, but instead depends on capacities that can become fatigued with repeated use or suboptimal physical conditions. Rather,

⁷ This model accounts for some of the empirical evidence that has put Baumeister’s model into doubt, such as the findings that belief in unlimited willpower enhances self-control, as does merely gargling with glucose (Molden et al. 2012; Job et al. 2010). In other words, willpower does not merely seem to be a physical resource that can be depleted and replenished with energy sources such as glucose.

it is almost always possible for us to exert willpower but significantly more difficult to do so in certain circumstances.

When we are in an unpleasant state and/or have effortfully exerted self-control recently, we become less motivated to engage in additional acts of self-control that would advance our long-term goals. For example, when we are hungry, we are motivated to alleviate the hunger by seeking out food or engaging in otherwise rewarding behavior. We become more sensitive to food-related cues, such as delicious aromas. Although we might have other important goals—such as attending to a lecture in order to get a good grade on a test—hunger reorients our attention and reduces our motivation to expend the cognitive effort required to engage with the lecture.

If you have ever had a stressful day at work after getting little sleep, you may have noticed that you are more inclined to eat junk food for dinner or lazily watch television instead of pursuing more productive activities. You likely realize it is not because you physically *could not* avoid eating junk food, but merely lacked sufficient desire to adhere to your diet in the face of a delicious, highly rewarding candy bar. Reductions in self-control impact our emotional reactions as well. In states of ego depletion, we lose our motivation to regulate our emotional reactions and may act more impulsively, especially in states that increase our sensitivity to negative emotions and impair our executive functioning. We *feel* that the agent who harmed us deserves to be treated in a certain way, strongly desire to act in accordance with that feeling, and do not want to expend the effort to suppress the reaction until we can fully evaluate the situation (e.g. yelling at your roommate before asking for an explanation). When we are ego depleted, sanctioning the agent is more rewarding than trying to treat them fairly by holding back.

I will analyze evidence that suggests that the internal states I have mentioned—hunger, sleep deprivation, and bad moods—impact our emotional responses to moral transgressions and our ability to control our behavior. For each negative state, I will explain how it affects our reactive attitudes and capacities to control our reaction and determine fitting sanctions. When I discuss particular components of our reactive attitudes—evaluative, bodily, and motivational—I am referring to the distinctions I provided earlier using the Attitudinal Account. These components can help us to gain a finer understanding of how these irrelevant influences can make our reactive attitudes go wrong.

4.1 Hunger

A study of parole decisions by Israeli judges suggests that hunger and fatigue may have an effect on their decision-making abilities. The study found that judges are more likely to make favorable rulings at the beginning of the day or after a food break; the probability of a favorable ruling declines steadily throughout the day, nearly hitting zero, until a food break brings the probability back up to around 65% (Danziger, Levav, & Avnaim-Pesso 2011). While the researchers did not have the ability to determine why the food break increased favorable outcomes, it is clear that certain factors irrelevant to the judicial decisions impacted the outcome. One potential explanation is that fatigue and hunger led the judges to rely more upon their emotional reactions than deliberate reasoning because they did not have the cognitive capacities to fully attend to the cases.

Assuming that deliberating on parole decisions in court is mentally taxing, judges should become progressively more fatigued throughout the day as they exert effort on cases. Additionally, feeling hungry could further hinder their ability to make parole decisions by making it more difficult to focus on the case and devote mental energy to making a decision. A

food break alleviates these negative states twofold: it eliminates the distraction of hunger and improves mood because eating is a pleasurable and rewarding activity. The willpower-boosting effect of merely tasting food is supported by a study that indicated that merely gargling with a sugar solution without swallowing bolstered self-control (Molden et al. 2012). So, even in the absence of hunger, tasty food can increase our motivation to focus on cognitively demanding tasks and suppress the impulses that result from our emotional reactions, such as punishing more severely. In other words, hunger may reduce one's ability to assess the evaluative component of their reactive attitude—the evaluation of what an agent deserves for their action—and may make it more difficult to refrain from acting on the motivational component.

It should be noted that the magnitude of the effect in Danziger et al.'s study has been called into question. Andreas Glöckner (2016) notes that a significant portion of the effect may be accounted for by non-random case ordering due to case complexity or presence of an attorney. However, Glöckner uses his research to conclude that the study does not provide *conclusive* evidence that the judges are impacted by factors extraneous to the case details. While deliberate case ordering provides an alternative explanation for the magnitude of the effect found, it does not explain away the entire effect. It may not be the case that hunger and fatigue *dramatically* impact judges' decisions, but there is still evidence that their decisions are impacted to some degree. Future research is needed to investigate this effect further. I will later discuss a more recent study that examines the effect of negative moods on judges' legal decision-making.

4.2 Sleep deprivation

A similar effect is seen when individuals are tired or sleep-deprived. Baumeister reports that “the self's resources are restored during sleep and then become progressively depleted during the day, especially insofar as the day makes demands for decision-making and self-

control. Hence failures of self-control are rare in the morning and become progressively more likely as the day wears on” (2002: 673). I believe this characterization is still viable under the motivational model of ego depletion. Exerting self-control throughout the day reduces our motivation to continue to exert self-control, and feeling fatigued, much like feeling hungry, can orient us towards more rewarding goals. However, sleep deprivation introduces a host of biological impairments that have not been shown to occur due to hunger. I will elaborate on these impairments to show how they might affect our reactive attitudes and sanctioning behavior.

A review on the effects of sleep deprivation by William Killgore (2010) highlights that one of the “most profound effects of sleep deprivation is an alteration in normal mood and emotional functioning, which may affect the assessment of risk and the types of judgments and decisions people ultimately make” (123). Sleep deprivation has been shown to impair complex moral reasoning (Olsen et al. 2010) and lead participants to violate the moral positions they adhere to when adequately rested (Killgore et al. 2007). These outcomes are not surprising given the large body of literature on the cognitive effects of sleep deprivation. Research suggests that lack of adequate sleep can increase our sensitivity to emotionally aversive stimuli (e.g. amplify the bodily component) and reduce our capacity to regulate our emotions and restrain our behavior.

Yoo et al. (2007) demonstrated that sleep deprived participants show significantly greater amygdala activation and reduced medial-prefrontal cortex connectivity in response to emotionally aversive stimuli than controls. These results suggest that sleep deprivation strengthens the bodily component of emotions (via greater amygdala activation) and reduces our control of the motivational component (via reduced prefrontal cortex connectivity). Across numerous studies and tasks, participants consistently display greater proneness to negative

emotions, longer perseverance of negative associations, reduced empathy, increased aggression and impulsivity in social situations, and reduced capacity for effective communication (Krizan & Hisler 2016; Harrison & Horne 2000).

These deficits are particularly relevant to moral evaluation. Lack of adequate sleep leads us to have stronger emotional reactions to moral transgressions and impairs the faculties we rely on to handle the situation fairly. Desire to alleviate the physical discomfort of sleep deprivation, by sleeping or pursuing other rewarding behavior, further alters our ability to control our emotional reactions. The aforementioned study by Killgore et al. (2007) provides evidence that sleep deprivation can lead to more severe punishments. The moral scenarios used in this study resemble the trolley problem, where participants are asked whether it is appropriate to perform a harmful action in order to save a greater number of lives (e.g. pulling a lever to reroute a train or pushing someone off a bridge to block the train). Participants answered a number of these dilemmas both at baseline and when sleep-deprived. The results indicate that participants are much more likely to endorse solutions that involve directing harming another individual (e.g. pushing someone off a bridge) when are sleep deprived, even though they would not typically consider these actions appropriate. Although this study did not investigate punishment, the results suggest that sleep deprivation makes participants less averse to personally harming others, even when no personal injury had been inflicted. From this, one may infer that sleep deprived individuals would be more permissive of inflicting harm as a sanctioning response to a personal moral injury, especially given their strengthened reactive attitudes and weakened impulse control.

4.3 Negative mood

Finally, incidental negative emotions have been found to impact wholly unrelated punishment decisions. Harlé and Sanfey (2007) found that incidental sadness, induced by a movie clip, significantly correlated with higher rejection rates of unfair offers in the ultimatum game. The ultimatum game (UG) is an economic game in which one individual (often a confederate) splits \$10 however they wish. If the participant accepts the offer, both individuals get paid. If the participant rejects the offer, neither individual gets paid. Because the participant purposely rejects free money to prevent the other individual from getting paid, the rejection constitutes a sort of punitive response. An increase in rejection rate suggests that incidental negative emotions can lead individuals to be more aggressive in social situations and more prone to making decisions influenced by their negative reactive attitudes.

Additionally, a study done by Eren and Mocan (2017) found that juvenile court judges were impacted when their undergraduate football team unexpectedly lost a game. In the week after their alma mater unexpectedly lost a football game, judges tended to impose longer sentences. Wins, expected losses, and games unrelated to their alma mater had no such impact, suggesting that their emotional investment in the team led them to be particularly sensitive to unanticipated negative results. These results suggest that, like the study of Israeli judges, even educated professionals are prone to the biasing effect of incidental negative emotions. More concerning, this effect was most pronounced in the sentencing of black juveniles by white judges, which suggests that the judges were most impacted in cases where they may have needed to compensate for their racial biases.⁸ Our experiences of unexpected bad fortune are not uncommon and certainly not limited to college sports upsets. If emotional upsets can affect the

⁸ If racial bias actually played a role, it is unclear whether the bias is implicit or explicit. Ideally, judges do not have explicit racial biases (even though implicit bias is obviously far from ideal). For this particular point, I do not think the distinction makes a significant difference.

decisions of trained professionals, it follows that these sorts of incidental negative emotions can also impact our sanctioning decisions. We may not even be aware that these unrelated events are impacting our decisions.

The explanation here is not dissimilar from those already given for hunger and sleep deprivation. When we are in a bad mood, even due to seemingly trivial causes like movies or sports, our motivation and attention shift toward behaviors that lessen the negative affect we are experiencing. In such a state, we are less able to regulate our emotional responses and behavior. In their analysis of self-regulation failure, Heatherton and Wagner (2011) identify negative moods as “[a]mong the most important triggers of self-regulation failure,” citing studies that link negative moods with aggressive behavior, various types of risky behavior, and failure to pursue long-term goals (2). These behaviors indicate emotional impulsivity and failure to consider long-term consequences, both of which apply to our reactive attitudes and sanctioning behavior.

From these studies, we can see that common and unexpected factors can impact the strength of our reactive attitudes. These effects span all three components of our reactive attitudes: the bodily component is strengthened, the motivational component becomes more difficult to control, and we have a reduced cognitive capacity to assess the accuracy of the evaluative component. The ego depletion caused by negative states makes it more difficult to prioritize our long-term goals—such as treating others fairly and maintaining our valuable relationships—over the more gratifying short-term goal of expressing our emotional reaction and sanctioning the elicitor. Further, certain states—such as sleep deprivation—introduce additional biological handicaps on our ability to regulate our emotions, control our behavior, and make well-informed sanctioning decisions. However, more research is needed to shed light on the

ways that other distinctive negative states impact our emotional regulation and decision-making capacities, especially in specifically moral contexts.

For example, it would be helpful to investigate the effects of sleep deprivation on moral sanctioning decisions, rather than vignettes that do not involve personal moral injury (Killgore 2007). Testing participants' emotional reactivity and willingness to punish in response to moral harm would give a better idea of how significant alterations in our executive capacities affect our motivational tendencies. Additionally, research on the effects of hunger on moral reactive emotions and sanctioning decisions would be helpful in interpreting the results of past studies (Danziger et al.) and provide valuable information on how hunger can affect our emotions and behavior.

The effects of these states can be summed up using the distinctions I drew from the Attitudinal Account. Negative states, through ego depletion or other biological handicaps, can amplify the bodily component of our reactive attitudes, such that their felt quality is stronger and more overpowering than it would be at baseline. The amplified bodily component makes it even more difficult and less desirable to control the motivational responses that it prepares the body to undertake. Because we are more overwhelmed by the emotion and less able to suppress our impulses, we are less likely to seek out additional evidence to confirm or disconfirm the evaluation that an agent has expressed ill will by violating a moral demand. For these reasons, we are more likely to sanction when it is wholly undeserved (because we did not uncover an extenuating explanation) or sanction more severely than the agent deserved (because our emotional response does not accurately track the level of ill will expressed).

4.4 Objections

I will now address a number of objections to the arguments I have made so far. First, one might object that our reactive attitudes may produce a more appropriate response when we do not have the cognitive faculties to restrain them. For example, in the UG study where ego depleted participants reject unfair offers more frequently than controls, one might say that this is the most appropriate response to being offered an unfair split. These participants are not deterred by their own desire to gain money and instead deliver the proposer their just deserts for offering such an unfair split. Because our default state involves some level of emotional regulation, our emotions are actually most fitting when we are not regulating them. Similarly, in other cases discussed above, such as the studies of judge's decisions, one might argue that the reactive attitudes are tracking others' bad behavior or expressions of ill will effectively, such that we overregulate them unless we have diminished motivation to do so because of hunger, sleep deprivation, or bad mood.

A few responses are possible here. In the UG study cited above, participants were confined to accepting or rejecting the offer and had no opportunity to express their emotions. Although rejection is an appropriate response to an unfair offer, it is more economically rational to accept the offer, no matter the amount. A study by Xiao & Houser (2005) highlights a design flaw in this study: they found that participants rejected unfair offers less frequently when they were allowed to express their negative emotional reaction to the proposer. By not allowing the participants to sanction by expressing their emotional response, it is unclear how significantly irrelevant factors—sleep deprivation or sadness—impacted their decisions. What we can discern from the studies is that watching a sad movie did lead participants to reject more frequently, suggesting that they had stronger emotional reactions and less inclination to restrain their

punitive impulses. By conducting additional studies where ego depleted subjects are allowed to express their emotions, we can better determine how significantly these factors impact emotional reactions and rational decision-making.

Another response concerns the fact that Strawson advances a backward-looking account of moral responsibility, where we sanction based on what an agent deserves. At baseline, we can assume that people try their best to determine what an agent deserves for their transgression by regulating their immediate impulses and rationally assessing the situation. However, ego depletion may hinder this process by interfering with all three components of our reactive attitudes. For one, if the bodily component of our reactive attitudes is stronger and less subject to modulation by the prefrontal cortex, then our emotions will feel more overwhelming and our capacities to evaluate the situation and restrain our behavioral impulses will be reduced.

It may be helpful to view each component of the reactive attitudes as providing feedback to the others. When you first experience a reactive attitude, you have perceived that you have been harmed and feel the extent of the injury, which motivates you to retaliate. Ideally, the bodily component of the reactive attitude provides you with valuable information on the severity of the ill will expressed in the injury and motivates an appropriate sanctioning response. Additionally, any updates in our evaluation—such as an agent's extenuating excuse—should alter the strength of the bodily component to reflect the new information. If the bodily component—the feeling—of the reactive attitude is particularly strong and overwhelming, you may be more inclined to take your evaluation (that you have been morally harmed) for granted and less inclined to suppress your behavioral impulses.

If someone wanted to argue that stronger emotional responses are more fitting, they must support the claim that we make more accurate sanctioning decisions in states of ego depletion

relative to baseline. However, we often regret the decisions we make in states of ego depletion because they are worse in respect to our long-term goals, such as not finishing work, eating cake while on a diet, or angrily insulting a partner. These are not merely cases where being more *emotional* is bad. Rather, our reactive attitudes *do* have a cognitive component; when this component is overwhelmed by stronger bodily and motivational components, the reactive attitude is less fitting than it is at baseline. Research shows that alcohol intoxication increases intentions for aggressive behavior, restricts attention to cues related to the conflict, and disinhibits aggressive behavior (Stappenbeck & Fromme 2014). These effects—impaired attention and ability to regulate emotions and behavior—are similar to those produced by ego depletion. If one wanted to claim that we overregulate our emotions at baseline, then they must concede that we react and sanction more accurately when we are intoxicated, given that intoxication bears important similarity to ego depletion.

Additionally, this study highlights that people often *choose* to regulate their emotions and behavior when they believe they have reason to (i.e. when they know they are intoxicated), even when experiencing high levels of emotional arousal. Emotional regulation is a process that, at least to some extent, we consciously engage in, especially when we have reason to believe that our emotional reactions are altered in some way. If people thought their bodily and motivational responses were more accurate when unregulated, it is unlikely that they would put forth the effort to regulate them. It would be much easier to accept that our unregulated emotions are more accurate and sanction in accordance with them than to attempt to regulate them first. It may be the case that awareness of the effects of ego depletion may make individuals less trusting of their emotional responses when ego depleted, just as they seem to be when knowingly intoxicated. For

this reason, it is important to make people aware of how their reactive attitudes may be affected by these irrelevant influences.

Under a forward-looking, consequentialist account of responsibility, this objection would pose no problem. Forward-looking accounts prioritize changing an agent's future behavior over considerations of what they deserve, so our emotions need not track any properties that would provide information about what the agent deserves. If we sanction only to prevent agents from performing moral transgressions again in the future, then it is likely beneficial to have stronger, more impulsive responses to their actions. From an evolutionary perspective, it makes sense that states of ego depletion would lead us to punish more impulsively. Food and sleep are necessary for survival, so when we lack these resources, it is advantageous to be primed to pursue them. These conditions lead us to sanction those who injure us more severely, which may deter them from injuring us again in the future. Neutralizing threats from other agents further increases in-group survival fitness. However, in modern conditions where maintaining good social relationships is both personally and legally valuable, we tend to evaluate actions based on the agent's intent (Aharoni & Fridlund 2011). This means that our ability to use our emotions to discern what an agent deserves is crucial.

To summarize, our moral reactive attitudes do not fittingly respond to an agent's wrongdoing when those attitudes vary depending on whether an individual has eaten, slept, or watched a sad movie. If the reactive attitudes are meant to track an agent's quality of will, then the inconsistency introduced by these irrelevant influences problematically prevents them from doing so. Due to this inconsistency, I argue that we should not allow our reactive attitudes to affect our treatment of an agent before we assure that these attitudes are fitting. In other words, we must make sure that our reactive attitudes are not subject to irrelevant influences before

allowing them to inform our sanctioning responses. The reactive attitudes we experience in the absence of irrelevant influences are valuable elements of our social relationships and we should strive to maintain their accuracy.

If we sanction an agent based on these adulterated reactive attitudes—through negative emotional reactions or more deliberate punishment decisions—then we risk unfairness in a number of ways. An unfitting reactive attitude may lead us to sanction when an agent does not truly deserve it or sanction more severely than the agent deserves. Further, if our reactive attitudes are altered due to a state of ego depletion—due to negative emotion, hunger, or tiredness—we may be unable to properly control our emotions or properly integrate them into our decision-making process.

Even in light of these inconsistencies, I do not believe that our reactive attitudes cannot usefully inform our sanctioning decisions. If we can prevent our reactive attitudes from immediately affecting an agent, they can later provide valuable information when we deliberate on an appropriate sanctioning response, if such a response is indeed warranted. Just as fear can alert us to the presence of danger, resentment calls our attention to the possibility of having been the victim of ill will and prepares us to respond with an appropriate sanction. A fitting reactive attitude, in combination with our rational deliberative capacities, can help us determine the severity of the ill will expressed. And the level of ill will expressed in an action is critical to determining what the agent deserves in response.

As Strawson pointed out, we cannot and should not eliminate our reactive attitudes and permanently assume the objective stance—I agree with him on this point (1962: 202-203). However, I do think that we can learn to control our reactive attitudes by practicing methods of self-control, a strategy that Strawson would likely find desirable given the research presented. He

does not “deny the possibility and desirability of redirection and modification of our human attitudes” in light of studies that suggest that they are “a prime realm of self-deception, of the ambiguous and the shady, of guilt-transference, unconscious sadism and the rest” (Strawson 203). Given this stance, I argue that Strawson would similarly endorse putting forth effort to reduce or eliminate the effect of common irrelevant influences on our reactive attitudes and sanctioning decisions. I will elaborate on this point in the following section. If we are to use reactive attitudes to inform our sanctioning decisions, we must determine how to gauge their fittingness and restrain them when they do not appear to meet the criteria.

5 CONTROLLING OUR REACTIVE ATTITUDES

So far in this paper, I have argued for the following:

1. If irrelevant influences sometimes prevent our reactive attitudes from accurately tracking an agent’s quality of will and reflecting how the agent deserves to be treated for their moral action, then we are not justified in allowing these attitudes to affect our treatment of the agent until we assure they have not been subject to irrelevant influence.
2. Evidence shows that incidental internal states can influence our moral reactive attitudes, such that they do not accurately track the quality of an agent’s will.
3. Therefore, we are not justified in allowing our reactive attitudes to affect our treatment of agents in response to their moral actions until we can make sure that our reactive attitudes are not altered by irrelevant influences.

In the previous section, I provided evidence that common internal states do significantly alter our reactive attitudes and affect our sanctioning decisions. When we are in these states, our reactive attitudes may be unfitting, in that they do not accurately correspond to an agent’s ill will.

Our unfitting reactive attitudes can be problematic in three ways: (1) They may lead us to sanction an agent before giving them a chance to offer an explanation, (2) they may lead us to forgo asking the agent for an explanation altogether, and (3) they may provide bad information that leads us to sanction more severely than is deserved.

The first two problems are not exclusive to unfitting reactive attitudes: our reactive attitudes may lead us to sanction in the absence of justification even with they do accurately correspond with the agent's ill will. In these cases, the problem is that we do not *know* if they are fitting. I take this to indicate that control over our reactive attitudes is crucial even when they are *apparently* fitting.⁹ We are not justified in imposing sanctions unless we guarantee that an agent's *apparent* ill will is *actual* ill will. However, these cases are still most problematic when the reactive attitude is unfitting, in that it does not even accurately correspond to the apparent ill will. Experiencing a reactive attitude in a state of ego depletion increases the chance of sanctioning before the agent explains their action. If their explanation renders the reactive attitude unfitting, then we would have imposed a wholly undeserved sanction. If they were actually expressing ill will, then we still have a greater chance of unfairly imposing a harsher sanction than they deserve. These cases highlight the problem of impulsivity in our reactive attitudes, which will likely be amplified when we are in states conducive to unfitting reactive attitudes.

Contrary to (1) and (2), (3) specifically deals with reactive attitudes that are unfitting due to irrelevant influences. Reactive attitudes are valuable insofar as they help us to gauge the degree of ill will expressed in the action and determine an appropriate response, just as fear allows us to gauge the dangerousness of a stimulus and act accordingly. When our reactive

⁹ Apparently fitting insofar as the reactive attitude accurately corresponds to the perceived level of ill will. If the agent has a viable excuse, we can no longer claim that the attitude was fitting because the ill will was merely illusory. But if they do not have an excuse, the reactive attitude fittingly captures their ill will.

attitudes are unfitting, they provide us with inaccurate information about the magnitude of ill will that can misinform our sanctioning decisions. For example, let's say that you are in court to protest a speeding ticket that you received. The ticket was administered on the basis of the officer's report, environmental stimuli such as traffic signs, and the device used by the officer to gauge your speed. In court, you provide your own report that you were not driving at the speed reported. If a test of the speedometer finds it to be inaccurate, the fine for the speeding ticket may be reduced or eliminated depending on the level of inaccuracy. Additionally, if the police officer admits that he was tired and thought you speeded in an active school-zone when school was not in session, it would be unfair for you to pay the fine for that violation. I argue that reactive attitudes should be treated similarly. Although they are not the only piece of evidence used to determine our sanctioning decisions, unfitting reactive attitudes have the potential to lead to sanctions that do not fit the violation.

In the speeding ticket examples, testing the accuracy of important equipment and evaluating the reliability of the officer's testimony corrected the inaccurate evidence. These errors were only corrected because the ticketed individual was allowed the chance to protest the ticket in court, where the accuracy of the evidence could be investigated. Again, I argue that the case is not dissimilar with our reactive attitudes. By allowing an individual to answer for their action by offering an explanation, we give them the chance to avoid an entirely unwarranted sanction. If their explanation does not exempt them, we must assure that we do not impose an inappropriate sanction by evaluating the accuracy of our evidence and assuring that we have the cognitive resources to integrate this information into an appropriate decision.

I will offer three criteria for using the reactive attitudes in our sanctioning decisions: (1) We cannot allow our reactive attitudes to affect our treatment of an agent before they have

explained the action, (2) we must be able to gauge the fittingness of our reactive attitudes, and (3) we must have the cognitive resources to integrate all of our evidence into an appropriate sanctioning decision. I believe the key in meeting these criteria lies in recognizing states of ego depletion and practicing methods of self-control. Because we can understand the emotional nature of the reactive attitudes using the Attitudinal Account, it becomes easier to see how implementing various methods of self-control could alleviate the effects of irrelevant influences. By recognizing the evaluative, bodily, and motivational components of our reactive attitudes, we can understand how each component can be controlled in order to avoid treating others unjustly.

The first criterion can be met by implementing a sort of if-then rule for one's behavior: if I think an individual has morally injured me, then I will make them explain their actions before subjecting them to a sanctioning response. Asking for an explanation without subjecting them to any negative treatment may be difficult, as our emotional expressions and vocal tone may be hard to suppress. However, these sorts of responses are not highly problematic because they do not impose a considerable amount of harm on the agent. If an agent did not genuinely express ill will, then your unfair reactive attitude will be quickly corrected and you can apologize for any undeserved harm. It is crucial to focus on more severe forms of sanctioning that have a greater potential to harm the agent and alter your relationship with them.

The second and third criterion may be met by learning to recognize cues that indicate ego depletion. We cannot guarantee that our reactive attitudes are free from irrelevant influence—at least due to the factors explicated previously—unless we can recognize signs that we are ego depleted. Learning the signs indicative of ego depletion can alert someone to the fact that they may need to subject their reactive attitudes to closer scrutiny to assure that they are not unfitting. Similarly, recognizing ego depletion may signal to someone that their cognitive capacities are

too taxed to fairly deliberate on a sanctioning decision, if such a decision is warranted. It has been suggested that individuals with strong emotion regulation skills are better at controlling their behavior when intoxicated than others are when sober. Researchers explained this result by suggesting that those high in emotion regulation skills feel that they must compensate for the effects of alcohol by engaging in self-regulation more readily (Stappenbeck and Fromme 2014). I argue that the same can be done with ego depletion; when someone is aware that they are ego depleted, they can consciously employ strategies to compensate for the detrimental effects.

Many of us already do this to some extent, by acknowledging that we may be grumpier after skipping a meal or that we need a good night's sleep before a cognitively taxing exam. By paying attention to how our emotional states and action tendencies differ when we are overly taxed, we can identify when we are more likely to react to moral transgressions (or any situation) unfittingly. Even if we are not aware of our altered executive capacities, we can try to take note of any stressors that contribute to reduced functionality. I will now discuss methods of emotion regulation that may be used to help fulfill all three criteria.

Although it may not be natural to initially suppress sanctioning urges to ask for an explanation, we can habituate ourselves into doing so. If you decide to go on a diet, it may at first be difficult to choose a healthy fruit over your favorite desserts. However, if you reflect on your ultimate goal—to become healthier—you may feel motivated to overcome your desires for the sugary desserts, such that you are able to successfully choose fruit instead. Over time, by repeatedly choosing fruit over sugary desserts, it becomes more habitual (and thus easier) to continue eating the fruit.

The same can be said for initially suppressing our reactive attitudes. By reflecting on your goal to treat others fairly, you may find it easier to muster the motivation to initially

suppress your reactive attitudes to make sure they are fitting until such a procedure becomes habitual. Thinking about times when you felt as if another sanctioned you unfairly may facilitate this process. Additionally, you may think about times when you unfairly harmed another. By redirecting your attention from the slight at hand to your ultimate goal to treat others fairly, you stand a better chance at controlling your reactive attitudes (Gross 2015). This strategy is known as attention deployment, where one directs their attention away from cues that reinforce the negative emotion. By focusing on scenarios where an unfitting reactive attitude caused harm, you distract yourself from the evaluation that you have been harmed. This redirection can weaken the bodily and motivational components, such that you are less likely to sanction the agent before asking for an explanation.

Similarly, one could also engage in cognitive appraisal, which is a deliberate attempt to change the trajectory of an emotion. Instead of acting off the initial appraisal that you have been maliciously harmed, you may reappraise the situation as an opportunity to demonstrate kindness and understanding. Alternatively, you may question your initial appraisal that the act was performed maliciously and think of explanations that would frame the act as accidental or unintentional. This strategy also aims to reorient the evaluation that supports the bodily and motivational components of the reactive attitude.

Of course, this process will be more difficult in states conducive to unfitting reactive attitudes. The irrelevant factors discussed in the previous section—hunger, sleep deprivation, and negative emotion—all contribute to ego depletion, a state in which we have more difficulty directing our motivation and attention to our long-term goals. Due to our reduced self-control and mental capacities, we experience negative emotions more strongly, act more impulsively, and cannot perform cognitively demanding tasks as well. This increased difficulty introduced by

ego depletion highlights the importance of making such a process habitual, as habituation reduces the cognitive resources necessary to perform a task and increases the probability of successfully performing the task in a state of ego depletion (Wood 2016). By challenging ourselves to redirect our attention or reinterpret the situation every time we have a reactive attitude, we can learn to do it more readily when we are more taxed than usual.

If we have requested an explanation from an agent and find her actions to genuinely express ill will, then the primary task is assuring that our reactive attitude fits the moral transgression. If we have some awareness of how our emotional reactions differ when in a state of ego depletion, then we are better equipped to identify when our reactive attitudes may not fit the situation. This strategy assumes that we can take the reactions we have in a state absent of significant ego depletion as baseline—an assumption that I do not take issue with here. Given the evidence that our decision-making capacities are impaired in states of ego depletion, habituation is also an appealing option here. If we identify that our reactive attitudes have the potential to provide us with inaccurate information and that we may not be capable of making a well-informed decision, we can form habits that prevent us from sanctioning in such a state. For example, we may redirect our attention away from the moral transgression and refrain from deliberating on a sanctioning response until we recover from the state of ego depletion.

In these situations, it may also be effective to engage in active suppression of the reactive attitude. For example, the bodily component of resentment includes a number of physiological changes, such as increased heart rate and adrenaline. Practicing deep breathing or visualizing a calming scene may help to lessen the physiological effects of the reactive attitude, thus making it easier to suppress motivational urges and think about the situation more clearly. If all else fails,

one may simply choose to remove themselves from the situation until feeling more replenished and calm enough to deliberate on an appropriate sanction.

In short, we can use self-control methods such as habituation, attention deployment, cognitive reappraisal, or suppression to help assure that our reactive attitudes do not lead us to sanction too hastily or too harshly. By learning to detect the differences in our bodily reactions and thought processes in states of ego depletion, we can learn to identify when our reactive attitudes may be unfairly influenced by factors irrelevant to the level of ill will expressed in an agent's action. These measures can help us to determine sanctions that are commensurate with the moral transgression at hand, such that we can avoid inflicting unwarranted harm on others. However, these techniques cannot be learned overnight; learning to recognize fluctuations in emotional responses takes time and experience, as does the sort of habituation that would benefit our moral decision-making process. Considering this, controlling our reactive attitudes should be considered a skill that we can continually hone.

One might say that all of these suggestions are commonsense and provide no novel insight into our sanctioning behavior; after all, most have heard the advice to count to ten when you are upset before dealing with the stressor. However, as with all home remedies, it is important to understand why they work and how effective they actually are. By understanding the neural and psychological mechanisms underlying our reactive attitudes and sanctioning decisions, we can better identify solutions and understand why they are effective and when they are not. For example, cranberry juice is commonly touted as a solution for recurrent urinary tract infections, but recent research indicates that cranberries do not successfully prevent these infections (Barbosa-Cesnik 2011). Failure to evaluate this folk remedy may have led people to continue using it instead of pursuing medical treatment, thereby leading to more severe medical

issues. The same can be said for our solutions for impulsive sanctioning: we may have heard of a number of solutions, but these solutions must be investigated and empirically validated so we do not risk sanctioning unfairly. Additionally, we have neither uncovered all of the irrelevant influences that may affect our reactive attitudes nor have identified all of the strategies that may help to eliminate these influences.

Given that Strawson was aware of psychological research that revealed the dark side of the reactive attitudes, I believe that it is crucial to pursue this research further to determine how significantly our attitudes are irrelevantly influenced, which irrelevant factors pose a threat to our reactive attitudes, and how this influence may be eliminated. Although we should not aim to eliminate our reactive attitudes, we should learn how to gauge when they are accurately tracking the magnitude of ill will expressed in the action so we do not inflict undeserved harm.

6 CONCLUSION

In this paper, I have argued that we are not justified in using our reactive attitudes to sanction others before we guarantee that said attitudes are fitting. Although Strawson and other Strawsonian scholars have readily identified the reactive attitudes as emotions, little work has been done on how the reactive attitudes affect our thoughts and behavior. In order to assess the viability of Strawson's position—that the reactive attitudes are an integral feature of human socialization that should not be eliminated—it is important to understand how the reactive attitudes may lead us to treat others unfairly. I have argued that psychological research on reactive attitudes and moral sanctioning gives more reason to question our natural tendencies than Strawson acknowledged. While the reactive attitudes are a natural response to the moral actions of others, that does not mean that we can overlook the unfairness introduced by irrelevant

influences such as hunger, sleep deprivation, and bad moods. I agree with Strawson that our reactive attitudes should not be eliminated completely, but do think that we must learn to control our reactive attitudes and lessen the impact of irrelevant influences in order to justifiably use them to hold others morally accountable.

I have attempted to provide a framework for using our reactive attitudes in a way that reduces the unfairness introduced by irrelevant influences. Under the Attitudinal Account of emotion, the reactive attitudes should correspond to the relevant evaluative properties in the object of the emotion. In the case of the reactive attitudes, the relevant evaluative property is the quality of will of other agents as expressed in their actions. If a reactive attitude does not correspond accurately to the quality of an agent's will, it is unfitting. Using psychological studies that demonstrate the unreliability of our emotional reactions and punishment decisions, I have shown that our reactive attitudes are often unfitting—ego depletion may make our reactive attitudes more overwhelming and difficult to control. Unfitting reactive attitudes provide inaccurate information about the severity of the agent's transgressions, making them problematic for a merit-based account of responsibility. However, by practicing various methods of emotion regulation and self-control, we may reduce the level of unfairness introduced by irrelevant influences.

It may be the case that Strawson underestimated how easily our reactive attitudes could be interfered with, but this does not mean his account must be abandoned. Assuring that our reactive attitudes and sanctioning decisions are not subject to irrelevant influence is critical to the viability of the Strawsonian position. When these irrelevant influences are controlled for, our reactive attitudes can provide valuable information about an agent's wrongdoing that could not

be apprehended by purely rational means. Our reactive attitudes, when fitting, help us to assess others morally and determine appropriate sanctions for moral wrongdoing.

I conclude that the Strawsonian account of moral responsibility can withstand the threats posed by irrelevant influences and provides important insight into our practices of holding others accountable for their actions. However, it is still imperative to conduct further psychological and neuroscientific research to gain a better understanding of our moral emotions and sanctioning decisions, to further reduce the risk of causing undue harm to others. Using the framework of emotion I suggested, it may be easier to discover and study new methods of emotion regulation that apply to our specifically moral emotions.

REFERENCES

- Aharoni, E. & Fridlund, A. J. (2011). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and the Law*, 18(4), 599-625.
- Barbosa-Cesnik, C., Brown, M., Buxton, M., Zhang, L., DeBusscher, J., & Forman, B. (2011). Cranberry Juice Fails to Prevent Recurrent Urinary Tract Infection: Results From a Randomized Placebo-Controlled Trial. *Clinical Infectious Diseases*, 52(1), 23-30.
- Baumeister, R. F. (2002). Yielding to Temptation: Self-Control Failure, Impulsive Purchasing, and Consumer Behavior. *Journal of Consumer Research*, 28(4), 670-676. doi:10.1086/338209
- Baumeister, R. F., Vohs K.D., Tice D.M. (2002). The Strength Model of Self-Control. *Current Directions in Psychological Science*, 16(6), 351-355.
- Beck, R., & Fernandez, E. (1998). Cognitive-Behavioral Therapy in the Treatment of Anger: A Meta-Analysis. *Cognitive Therapy and Research*, 22(1), 63-74.
- Eren, O., & Mocan, N. (2017). Emotional Judges and Unlucky Juveniles. *SSRN Electronic Journal*. doi:10.3386/w22611
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist. *Proceedings of the National Academy of Sciences*.
- D'arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research*, 61(1). doi:10.2307/2653403
- Deonna, J., and Teroni, F. (2012) *The Emotions: A Philosophical Introduction*. London: Routledge.
- Glöckner, A. (2011). The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated. *Judgment and Decision Making*, 11(6), 601-610.
- Gross, J. J. (2015). Emotion Regulation: Conceptual and Empirical Foundations. In *Handbook of emotion regulation*. New York: The Guilford Press.
- Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences*, 15(3), 132-139. doi:10.1016/j.tics.2010.12.005
- Harlé, K., & Sanfey, A. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion*, 7(4), 876-881.
- Harrison, Y., & Horne, J.A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, 6(3), 236-249.

Inzlicht, M., & Schmeichel, B. (2012). What is Ego Depletion? Toward a Mechanistic Revision of the Resource Model of Self-Control. *Perspectives on Psychological Science*, 7(5), 450-463.

Job, V., Dweck, C., & Walton, G. (2010). Ego depletion: Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science*, 21, 1686–1693.

Killgore, W., Killgore, D., Day, L., Li, C., Kamimori, G., & Balkin, T. (2007). The Effects of 53 Hours of Sleep Deprivation on Moral Judgment. *Sleep*, 30(3), 345-352.
doi:10.1093/sleep/30.3.345

Killgore, W. D. (2010). Effects of sleep deprivation on cognition. *Progress in Brain Research*, 105-129. doi:10.1016/b978-0-444-53702-7.00007-5

Krizan, Z., & Histler, G. (2016). The Essential Role of Sleep in Self-Regulation. In K. D. Vohs & R. F. Baumeister (Authors), *Handbook of self-regulation: research, theory, and applications*. New York: The Guilford Press.

Lewis, C. S. (1987). The Humanitarian Theory of Punishment. *Issues in Religion and Psychotherapy*: 13(3).

Mocan, N. H., & Eren, O. (2016). Emotional Judges and Unlucky Juveniles. *SSRN Electronic Journal*. doi:10.2139/ssrn.2914649

Molden, D. C., Ming Hui, C., Scholer, A. A., Meier, B. P., Noreen, E. E., D'Agostino, P. R., & Martin, V. (in press). The motivational versus metabolic effects of carbohydrates on self-control. *Psychological Science*.

Olsen, O. K., Pallesen, S., & Jarle, E. (2010). The Impact of Partial Sleep Deprivation on Moral Reasoning in Military Officers. *Sleep*, 33(8), 1086-1090. doi:10.1093/sleep/33.8.1086

Stappenbeck, C. A., & Fromme, K. (2014). The effects of alcohol, emotion regulation, and emotional arousal on the dating aggression intentions of men and women. *Psychology of Addictive Behaviors*, 28(1), 10-19. doi:10.1037/a0032204

Stich, S. (1978). Beliefs and Subdoxastic States. *Philosophy of Science*, 45(4), 499-518.

Strawson, P. F. 1962. "Freedom and Resentment." In *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Ed. by Laura Waddell Ekstrom. Boulder, CO: Westview, 2001. 183-203.

Teroni, F., & Deonna, J. (2012). *The Emotions. ; A Philosophical Introduction*. Routledge.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, 24(2), 227-248.
doi:10.5840/philtopics199624222

Wood, W. (2016). The Role of Habits in Self-Control . In K. D. Vohs & R. F. Baumeister (Authors), *Handbook of self-regulation: research, theory, and applications*. New York: The Guilford Press.

Yoo, S. S., Gujar, N., Hu, P., Jolesz, F. A., & Walker, M. P. (2007). The human emotional brain without sleep—a pre-frontal amygdala disconnect. *Current Biology*, 17, R877–R878.