Journal of Cognition and Neuroethics

Exploring the Status of Free Will in a Deterministic World: A Case Study

Catherine Gee
University of Waterloo

Biography

Catherine Gee is a PhD student in philosophy at the University of Waterloo in Waterloo, Ontario. Her primary research interests lie at the intersection of philosophy and psychology, and in philosophy of psychiatry in particular. Issues concerning the proper classification of mental disorders and their implications for treatment are one of the current topics she is working on, in addition to projects in philosophy of mind and philosophy of science.

Publication Details

Journal of Cognition and Neuroethics (ISSN: 2166-5087). March, 2015. Volume 3, Issue 1.

Citation

Gee, Catherine. 2015. "Exploring the Status of Free Will in a Deterministic World: A Case Study." *Journal of Cognition and Neuroethics* 3 (1): 175–194.

Exploring the Status of Free Will in a Deterministic World: A Case Study

Catherine Gee

Abstract
Developments in sciences have been uncovering causal mechanisms which have improved our understanding of human character and behaviour, which seem to be the result of deterministic forces. This leads to questions regarding the status of free will as it is often argued to be incompatible with determinism. Individuals with anorexia nervosa present a unique case study that can be used to apply the philosophical arguments involved in the free will debate to real agents whose illness may, at least partly, be the result of deterministic causal mechanisms. The agent has little to no control over most of these causal mechanisms, which seems to imply that she cannot have free will. However, I will attempt to argue that this is not necessarily the case. While the anorexic agent cannot control the causal mechanisms which contribute to her illness, she is able to retain control over her intentional actions. She possesses the capacity to reflect critically on her first-order preferences and desires and either identifies with these preferences or changes them via her higher-order desires and acts accordingly. As such, the power and choice are ultimately her own.

Keywords
Free will, determinism, compatibilism, anorexia nervosa

Developments in sciences such as biology, neuroscience, psychology, and psychiatry have been uncovering causal mechanisms which have improved our understanding of human character and behaviour, which seem to be the result of deterministic forces. This leads to questions regarding the status of free will as it is often regarded as incompatible with determinism. While the free will versus determinism debate has always been a favorite in philosophy, most of it is focused on the philosophical arguments themselves and not on how the philosophical positions actually apply in real-world cases. Many of the examples in the philosophical literature are thought experiments or fictional cases used to illustrate an argument, when it would be more useful to utilize a real-life example to see how these philosophical arguments pan out when they are applied. The free will debate gives us a unique opportunity to do just this, as psychological research supports philosophy's deterministic position by providing empirical evidence regarding human behaviour that bolsters its claims. One such area of research examines the causal mechanisms that contribute to anorexia nervosa. By using anorexic patients as a case study to examine the philosophical arguments which attempt to reconcile free will with determinism, we are able to examine agents who are practically useful and

philosophically interesting to study. Individuals with anorexia nervosa are influenced by real causal forces which present a philosophical challenge for one who wishes, as I do, to find a way to argue that despite these forces, the individuals with anorexia are still able to retain free will.

This paper will first present the philosophical position in support of a deterministic perspective and show how the empirical research on anorexia in psychology supports this philosophical claim. In the second section, philosophical attempts to reconcile free will with determinism (the compatibilist position) will be presented and shown why they are inadequate when applied to the real-life anorexic agent. Finally, the third section will explore a solution that considers the role autonomy plays in free will. It will be argued that the anorexic agent is indeed able to retain some free will despite the deterministic causes of her illness, for while she may not be able to control the causes which contribute to her illness, she is still able to retain power and control over her intentional actions. She is able to do this either by identifying with her predetermined preferences and permitting the behaviour or action to commence (for example, refusing to eat), or she may instead opt for another course of action that is of her choosing (such as deciding to eat). The power and choice is ultimately her own, and as such, her responsibility.

1. Why Determinism?

A Philosophical Perspective

Robert Kane gives a wonderful overview of the literature on the free will versus determinism debate in his introduction to the *Oxford Handbook of Free Will*, which I will draw from to set up my project. To begin the discussion one must note that modern debates on this topic stem from two questions: (1) ***the Determinism Question*** which asks "Is determinism true?" and (2) ***the Compatibility Question*** which wonders "Is free will compatible (or incompatible) with determinism?" (Kane 2009, 2). Kane explains that the "[a]nswers to these questions give rise to two major divisions in contemporary free will debates" between determinists and indeterminists for the first question, and compatibilists and incompatibilists for the second (2009, 2). Regarding the determinism question, the prevailing (though certainly not uncontested) view in modern quantum physics is that we live in an indeterministic world. Quantum theory "denies that elementary particles composing the "system of the world" have exact positions and momenta that could be simultaneously known by any such intelligence (the Heisenberg Uncertainty Principle); and it implies that much of the behavior of elementary particles ...

is not precisely predictable and can by explained only by probabilistic, not deterministic, laws" (Kane 2009, 2). "[T]he uncertainty and indeterminacy of the quantum world", continues Kane, "is not merely due to our limitations as knowers but to the nature of the physical world itself" (2009, 2).

However, while universal indeterminism is the predominate view in physics, deterministic views regarding human behaviour have been on the rise in other sciences such as biology, neuroscience, psychology, and psychiatry (Kane 2009, 3). There are many reasons for this trend, but the most relevant one for our present purpose is that empirical research in human behaviour and action has given us "greatly enhanced knowledge of the influence of genetics and heredity upon human behavior ... greater awareness of biochemical influences on the brain; the susceptibility of human moods and behaviour to drugs... influences of psychological, social, and cultural conditioning upon upbringing and subsequent behaviour, and so on" (Kane 2009, 3). These findings seem to be pointing more and more towards our behaviour being "determined by causes unknown [unconscious] to us and beyond our control" (Kane 2009, 3). As a result, as Sam Harris argues:

> Today, the only philosophically respectable way to endorse free will is to be a compatibilist – because we know that determinism, in every sense relevant to human behavior, is true. Unconscious neural events determine our thoughts and actions – and are themselves determined by prior causes of which we are subjectively unaware. However, the "free will" that compatibilists defend is not the free will that most people feel they have. (2012, 16)

We will get to the compatibilist position in the next section, but first we will look at what Harris means by 'free will that most people feel they have,' which is typically referred to as Libertarian Free Will.

This concept can be seen as the "freest" of the free will positions as it asserts free will exists and the universe is indeterministic. Contemporary free will libertarians must be able to successfully deny determinism as well as deny the compatibility of free will and determinism (Kane 2009, 8). Unlike the issue compatibilists face with reconciling free will with determinism (again, more on that shortly), the problem for the libertarian is how to make sense of a conception of free will that is *incompatible* with determinism by finding a way that we can have free will in an indeterministic world (Kane 2009, 8). The problem with indeterminism is that an event may or may not occur, it is a matter of chance and, as Kane explains, "chance events are not under the control of anything, hence not under the

control of the agent. How then could they be free and responsible actions?" (2009, 8). Instead of assisting the libertarian position by escaping the problems with determinism, indeterminacy may undermine freedom rather than enhance it (2009, 8).

Libertarianism was a worthwhile attempt, for as Saul Smilansky explains, "it was supposed to allow a deep moral connection between a given act and the person, and yet not fall into being merely an unfolding of the arbitrarily given, whether determined or random" (2009, 1-2). However, Smilansky concludes that this project is not plausible because, as Adina Roskies states "[t]his view does not seem to cohere with any scientific picture that we know" (2006, 420). This is the general philosophical consensus on libertarian free will, though there are certainly still philosophers working on the libertarian view, including Kane. Because psychological research supports a deterministic view of human behaviour and action, and libertarian free will goes against empirical studies on how we make our decisions, we will put aside the libertarian view and turn to the psychological research that supports the philosophical case for determinism.

A Psychological Perspective

While the exact etiology of anorexia nervosa is not known, the psychological research is uncovering some important findings that are making promising headway in uncovering the causal mechanisms that underlie this eating disorder. In this section I will present some examples from the research on the deterministic causal forces that contribute to anorexia. Before I do so the diagnostic criteria for anorexia nervosa will be outlined to give an overview of the sort of characteristics the illness exhibits.

Anorexia Nervosa: DSM 5 Diagnostic Criteria

1. Restriction of energy intake relative to requirements, leading to a significantly low body weight in the context of age, sex, developmental trajectory, and physical health. *Significantly low weight* is defined as a weight that is less than minimally normal or, for children and adolescents, less than that minimally expected.
2. Intense fear of gaining weight or of becoming fat, or persistent behavior that interferes with weight gain, even though at a significantly low weight.
3. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or persistent lack of recognition of the seriousness of low body weight.

Restricting Type

During the last 3 months, the individual has not engaged in recurrent episodes of binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas). This subtype describes presentations in which weight loss is accomplished primarily through dieting, fasting, and/or excessive exercise.

Research Findings

A well cited study by Goss and Gilbert (2002) suggests that shame may play a large role in the anorexic's eating disorder. The researchers assert the restrictive symptoms may function in a shame-pride cycle where "feelings of internal and external shame lead to restriction and the subsequent weight loss (successful restriction) leads to feelings of pride so a shame-pride cycle develops where shame negatively reinforces and pride positively reinforces the primary symptoms of restriction and weight loss" (Troop et al. 2008, 481). Shame is considered to be an emotion that has a highly social component for it is concerned with a fear or anticipation of eliciting disgust in others, either in the presence of a real or imagined audience (Troop et al. 2008, 480). It has strong implications with eating disorders and anorexia is more commonly associated with external shame, and bulimia tied to internal shame. External shame is the result of a person's perception that others view her in a negative manner, in that "the self is perceived by others as an object of scorn, ridicule and contempt" (Troop & Redshaw 2012, 373). Internal shame is a negative reflection of how one views oneself such as "worthless, flawed, morally defective or unattractive" (Troop & Redshaw 2012, 373). The difference between the two can be summarized as "the experience of being shamed versus feeling ashamed" (Gilbert 1998, as cited in Troop et al. 2008, 481). A proneness to experiencing shame is positively related to issues with eating in female students, and women who exhibit symptoms of an eating disorder experience more guilt and shame regarding eating than do depressed individuals and student controls (Troop et al. 2008, 481). While these findings were among non-clinical female samples where it was not established whether eating disorders were present or not, this research gives a good point of comparison as currently ill and recovered sufferers of eating disorders reported higher levels of shame (related to character and eating) than a student comparison group (Troop et al. 2008, 481). Additionally, the current sufferers of eating disorders report higher levels of shame regarding their character and eating than those who had recovered (Troop et al. 2008, 48). One study found that inpatients with anorexia and bulimia reported higher levels of

internalized shame than patients with depression and anxiety diagnoses (Grabhorn et al. 2006, as cited in Troop et al. 2008, 481).

Another causal factor that may contribute to anorexia nervosa is a body image dysfunction. There are two main types of body image dysfunctions that have been identified, a perceptual body-size distortion and cognitive-evaluative dissatisfaction. The former occurs "when a person has difficulty accurately gauging her body size", such as when the anorexic estimates her body size is larger than it actually is (Cash & Deagle 1997, 108). The latter is concerning one's attitudes about her body image and is often referred to as "body dissatisfaction" or "disparagement" (Cash & Deagle 1997, 108). In this case the anorexic may have an accurate view of her body size, but is quite dissatisfied with her body's size or shape. These two types of body image dysfunctions seem to be largely independent (Cash & Deagle 1997, 108). While women without an eating disorder can more or less accurately perceive their bodies, patients with anorexia overestimate their body fat (Benninghoven et al. 2007, 55). They also have been demonstrated to be less satisfied with their current body shape than the control subjects and have a thinner ideal body size in relation to what they see as their perceived body size (Mohr et al. 2009, 1524). Body satisfaction is calculated by asking subjects to select an image of a body type that matches their ideal body size, and another matching their actual body size. The discrepancy between the subjects' self-perception and their ideal perception is interpreted as a measure of body satisfaction (Mohr et al. 2009, 1521). When conducting a satisfaction rating with functional magnetic resonance imaging (fMRI) for thinner self-images, subjects with anorexia demonstrated stronger activation of the insula and the lateral anterior prefrontal cortex, while the controls "showed a stronger recruitment of precuneus during body size estimation for fatter images" (Mohr et al. 2009, 1524). Mohr et al. conclude from this that they "were able to separate the two different dimensions of body image behaviorally and find neural correlates with different task-specific involvements for anorectic patients and healthy controls" (2009, 1524). Furthermore, the researchers were able to find that anorexic patients had a higher activation of the left insula for the thin condition in the satisfaction rating in contrast to the thin condition in the body size estimation task (Mohr et al. 2009, 1525-1526). Results in brain areas and body image studies "indicate alterations in the activation of posterior parietal regions for patients with anorexia nervosa, possibly related to spatial components of the body image" (Mohr et al. 2009, 1520). Mohr et al. suggest that their findings could imply this insula activity could be associated with "a stronger experience of emotions by the patient group while viewing the self-images" (2009, 1527). Additionally, "[i]t could be speculated that the higher activity of the insula for thin self-images in the anorectic group is a

consequence of higher emotional valence of the thinner self-images" (Mohr et al. 2009, 1527) as insula activation has been tied to emotional and interoceptive awareness (Mohr et al. 2009, 1526).

Finally, the dysregulation of reward-processing mechanisms is believed to play a central role in eating disorders by contributing to and maintaining the core symptoms (Alonso-Alonso 2013, 1082). There appears to be a specific dysfunction with the suppression of the desire to eat ("wanting" food) while the capacity to evaluate ("liking") food is preserved (Berridge 2009, as cited in Alonso-Alonso 2013, 1082). A recent study by Frank et al. utilized voxel-based morphometry (VBM), a technique for automatic computational neuroanatomy, to examine the difference in brain structure in patients with or who had recently recovered from anorexia, patients with bulimia, and healthy control women (Alonso-Alonso 2013, 1082). The researchers found the anorexic and bulimic patients had larger gray matter volume in the gyrus rectus/medial orbitofrontal cortex, an area of the ventral system, compared to the controls (Alonso-Alonso 2013, 1083). The gray matter volume in this area correlated with pleasantness ratings ("liking") during a sucrose taste perception test prior to scanning (Alonso-Alonso 2013, 1083). The anorexic subjects "had increased gray matter volume in the anterior insula in the right hemisphere" while the bulimic subjects had a similar effect on the left side (Alonso-Alonso 2013, 1083). Furthermore, the recovered anorexic subjects and those with bulimia had reduced gray matter volume in the dorsal striatum, an area associated with a measure of reward sensitivity ("wanting") (Alonso-Alonso 2013, 1083). These research results find the gyrus rectus/medial orbitofrontal cortex as a neural substrate common to both anorexia and bulimia nervosa, and since this area is predominately involved in the processing of pleasant stimuli (Gabenhorst & Rolls 2011, as cited in Alonso-Alonso 2013, 1083), "this opens the possibility that brain changes determining an elevated capacity to experience pleasure from food (liking the taste) could represent neurodevelopmental contributors to eating disorders and act as an initial trigger of compensatory behaviors (such as decreased wanting for food rewards) early on in the natural history" (Alonso-Alonso 2013, 1083-1084).

These studies only present a very small sample of the research on anorexia nervosa but are intended to demonstrate the various mechanisms that may contribute to this eating disorder. Heightened emotional responses such as shame or a dysfunctional body image, or physiological issues such as a dysregulation in the reward centers of the brain are all possible deterministic factors over which the anorexic agent has no control. Her refusal to eat, and thus maintain a healthy body weight, is a much more complicated issue

than her simply declining to eat. This empirical data combined with the philosophical argumentation for determinism now directs us to the question of free will.

2. Pursuing Compatibilism

Having affirmed the Determinism Question[1] by confirming the causation of anorexia is indeed deterministic, I will now turn to the Compatibility Question [2] which is the true purpose of this paper – if the anorexic's illness is the result of deterministic processes, can she somehow maintain her free will? Sam Harris argues she cannot, for "[h]ow can we be "free" as conscious agents if everything that we consciously intend is caused by events in our brain that we *do not* intend and of which we are entirely unaware?" (2012, 25-26). Free will can be seen as incompatible with determinism because "[t]o say that "my brain" decided to think or act in a particular way, whether consciously or not, and that this is the basis for my freedom, is to ignore the very source of our belief in free will: the feeling of *conscious* agency. People *feel* that they are the authors of their thoughts and actions, and this is the only reason why there seems to be a problem of free will worth talking about" (Harris 2012, 26). However, the Compatibility Question does not *assume* that if determinism is true then we necessarily lack free will, and the burden of proof lies with those who assert the two are incompatible (Kane 2009, 3). The question does, however, seem to *imply* incompatibility (and historically this was the assumption) for when we speak of the ability to make a decision we assume we are able to choose from a variety of different possibilities (Kane 2009, 4). Modern arguments for incompatibilism often stem from this assumption, "the requirement that an agent acted freely, or of his or her own free will, only if the agent had *alternative possibilities*, or *could have done otherwise*" (Kane 2009, 4). Kane (2009, 4) refers to this as the **Alternative Possibilities (AP) Condition** which is as follows:

> The case for incompatibility from this Alternative Possibilities Condition has the following premises:
>
> 1. The existence of alternative possibilities (or the agent's power to do otherwise) is a necessary condition for acting freely (of one's free will).

1. Is determinism true?

2. Is free will compatible (or incompatible) with determinism?

2. Determinism is not compatible with alternative possibilities (it precludes the power to do otherwise).

A modern, and widely discussed, argument in support of premise 2 is the Consequence Argument, which is informally stated by van Inwagen (1983) as follows:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (as cited in Kane 2009, 4)

"Up to us" is the wording that Kane also favors, for we feel that it is "up to us" "what we choose and how we act" (Kane 2009, 1). Just as we could not choose events in the past we also cannot choose the impact these past actions will have on our present and future acts. It thus appears we cannot do anything to alter our present actions as we do not have the power to do otherwise, and hence lack alternative possibilities (Kane 2009, 4). In order to hold a compatibilist position one must successfully defeat the Consequences Argument, in addition to other incompatibilist arguments. We will now turn to the compatibilist position.

Classical compatibilists often define freedom in terms of one having the power or ability to act. It then follows that in order to be free one must have (1) the power and ability to act as one chooses or desires which requires (2) an "absence of constraints or impediments" which prevent one from doing as he wills, desires, or chooses to do (Kane 2009, 4). Traditional examples of constraints or impediments include things like "physical restraints, lack of opportunity, duress or coercion, physical and mental impairment" and so on (Kane 2009, 4). Kane notes "[y]ou lack the freedom to meet a friend in a cafe across town if you are tied to a chair, are in a jail cell, lack transportation, someone is holding a gun to your head, or you are paralyzed" (2009, 4-5).

However, for our purposes here we are more interested in subtle, biological constraints or impediments such as dysfunctional reward processing mechanisms that prevent the anorexic from making a truly free choice. In order for an agent to have free will that is compatible with determinism the agent must still be able to choose despite deterministic conditions. For our anorexia case, what would such a choice look like? Benjamin Libet has an interesting reply that focuses on the role that consciousness plays in voluntary action. Libet conducted an experiment where the subjects flicked their wrists "at any time they felt the urge or wish to do so" and were able to perform this flicking

of the wrist free of any external limitations or restrictions (for example, without time limitations or specific timed intervals) (Libet 2009, 1). He discovered that the readiness potential, which is "a specific electrical charge in the brain," began 550 milliseconds before the flicking of the wrist (2009, 1). The subjects were not aware of the intention to act until 350-400 milliseconds – which is *after* the readiness potential starts, but their conscious awareness occurs 220 milliseconds *before* the motor action (Libet 2009, 1). So while the volitional process is initiated unconsciously, the conscious function can still, according to Libet, control the outcome (whether to flick one's wrist or not) (Libet 2009, 1). Libet concludes that this means free will is not excluded as conscious will appears around 150 milliseconds before the muscle is activated (even though it does follow the unconscious onset of the readiness potential) (2009, 2). This unconscious onset is not problematic for Libet, as we still have enough time in the 150 milliseconds to either allow or veto the muscle action. Libet states he was "able to show that subjects could veto an act planned for performance at a prearranged time. They were able to exert the veto within the interval of 100 to 200 msec, before the preset time to act" (Libet et al., 1983b, as cited in Libet 2009, 3). He continues that "[a] large RP [readiness potential] preceded the veto, signifying that the subject was indeed *preparing* to act, even though the action was aborted by the subject" (2009, 3). Libet argues that "the conscious veto may *not* require or be the direct result of preceding unconscious processes. The conscious veto is a *control* function, different from simply becoming aware of the wish to act" (2009, 4).

Philosophically, this certainly seems like a promising compatibilist position. The veto power successfully defeats incompatibilism via the Alternative Possibilities Condition by defeating the second premise:

1. The existence of alternative possibilities (or the agent's power to do otherwise) is a necessary condition for acting freely (of one's free will).
2. ~~Determinism is not compatible with alternative possibilities (it precludes the power to do otherwise).~~

However, psychologically the veto is still problematic as we are left having to explain the anorexic who is able to veto her desire to refuse food and can eat (thus can overcome her illness) versus the one who decides not to stop the desire to refuse food and isn't able to gain weight (and in turn, get better). Why is one successful with the veto power and the other is not? Does the anorexic who ultimately refuses her veto power want to encourage her illness, or does she just have a weaker veto (willpower) than the anorexic who is able to get better? Either way, Libet's account leaves these questions unanswered

and is thus unsatisfactory for our purposes. If we are to grant the anorexic agent some control over her illness we cannot do it via Libet's account and need to look elsewhere.

Smilansky presents a more radical solution to the problem of free will in a deterministic world. He finds libertarian free will impossible but also finds too many problems with compatibilism to go that route. He argues "if there is no libertarian free will, no one can be ultimately in control, ultimately responsible, for this self and its determinations…[i]f people lack libertarian free will, their identity and actions flow from circumstances beyond their control" (2009, 2). It is due to this lack of control that the compatibilist has to accept "a shallower sort of meaning and justification" for moral worth, and it is due to this shallowness and "complacent compliance with the injustice of not acknowledging lack of fairness and desert" that Smilansky disagrees with compatibilism (2009, 2).

While Smilansky is more in favor of hard determinism which denies free will, he does not fully endorse it either. The hard determinist would assert that responsibility is a non-issue for we deserve neither praise nor blame for our actions, as ultimately they are beyond our control. However, Smilansky argues that rejecting notions of responsibility makes the hard determinist "morally blind and a danger to the conditions for a civilized, sensitive moral environment" (2009, 4). Instead what Smilansky wants is a Community of Responsibility that is comprised of members whose choices determine the moral attitudes they receive (2009, 3). Smilansky argues that the assumption that one must either affirm compatibilism or incompatibilism is incorrect, and states that while the two are logically inconsistent this does not mean one cannot hold an intermediate and mixed position between the two (2009, 2). By rejecting the assumption that only compatibilism or incompatibilism can be correct we are able to stay closer "to the deepest issues on the free will issue" and "proceed along a new path that ultimately runs closer to the intuitive field than do either of the conventional monisms" (Smilansky 2009, 2). The rejection of the Assumption of Monism paves the way for the Community of Responsibility, which is achieved via Illusionism. This is "the position that illusion often has a *large and positive role* to play in the issue of free will" for we are "fortunately" deceived into thinking we have free will and this allows us to maintain "civilized morality and personal value" (Smilansky 2009, 5).

Smilansky is not suggesting we induce illusionary beliefs regarding free will, or maintain ones that we "fully realize" are illusionary (2009, 4). Instead we are to permit and humour the illusionary beliefs that are already in place, for they are doing more good than harm. "The sense of "illusion" that I am using", explains Smilansky, "combines the falsity of a belief with some motivated role in forming and maintaining that belief

– as in standard cases of wishful thinking or self-deception" (2009, 5). Smilansky fully acknowledges that these beliefs are false, for "[a]ll our actions, however an internalized and complex a form they make take, are the result of what we are, ultimately beyond our control" (2009, 6). However, because compatibilism and hard determinism, for Smilansky, are too problematic to endorse on their own, illusionism is the only solution to maintain some sense of moral order and responsibility in this deterministic world.

Smilansky's illusionism is bound to be highly criticized, and for good reason. His motivation for a Community of Responsibility is understandable as it aligns with the moral intuition that agents deserve praise or blame for their actions. However, while we all surely entertain false but useful beliefs in our everyday lives – "Our troubles will soon end, for tomorrow will be better than today" etc. – actively encouraging them as a philosophical position is too problematic to endorse. The problem is illusionism is essentially a cop out, a last-ditch effort to salvage free will if only in name. Is the prospect of denying free will so devastating to us as agents that we must retain some semblance of free will, even in name only? Sam Harris doesn't think so; in fact he asserts that instead of becoming fatalistic as a result of denying free will it has actually increased his feeling of freedom. He states "[m]y hopes, fears, and neuroses seem less personal and indelible. There is no telling how much I may change in the future" (Harris 2012, 46). One need not draw lasting conclusions about himself based on how he thought or behaved in the past, for "[a] creative change of inputs to the system – learning new skills, forming new relationships, adopting new habits of attention – may radically transform one's life" (Harris 2012, 46). This all sounds quite promising, but we have to remember that if we deny free will these changes of inputs to our system would have to be predetermined as well and I am not ready to throw in the towel just yet. We will look at one more account before we call it quits for the day.

Alfred Mele's discussion of autonomy and self-control may help provide additional insight to the task at hand, for he uses autonomy as an extension of free will. In order for one to be autonomous they must be able to choose or act freely, to some degree. Furthermore, they must have some level of self-control, which Mele states agents possess when they "have significant motivation to conduct themselves as they judge best and a robust capacity to do what it takes so to conduct themselves in the face of (actual or anticipated) motivation" (Mele 1995, as cited in Mele 2009, 2). However, even if one is an "ideally self-controlled" person and is thus able to manifest "perfect self-control"[3] Mele

3. Mele states such a person would achieve the four dimensions for perfect self-control: range, object, frequency, and effectiveness. For more detail see Mele 1995.

argues this still insufficient for autonomy. Mele uses Gerald Dworkin's explanation of autonomy to explain why, which is as follows:

> [A]utonomy is a second-order capacity to reflect critically upon one's first-order preferences and desires, and the ability either to identify with these or to change them in light of higher-order preferences and values. (Dworkin 1988, as cited in Mele 2009, 3)

Mele argues "[a]n ideally self-controlled person has this capacity and ability. However, even ideal self-control – no matter how frequently and successfully exercised – might not suffice for autonomy. If, as it seems, every process of critical reflection is regulated or guided by principles or values already in place, some principle or value will be presupposed or taken for granted in each process" (2009, 3). However, he does not think that this necessitates that we are unable to have any control if the world is deterministic. To assert "the thesis that there is at any instant exactly one physically possible future"[4] does not mean that there cannot be "*more than one physically possible future*" as causation does not depend, argues Mele, on the absence of these physically possible futures (2009, 3, my italics). Just because there is only one outcome does not necessitate that there were not any other possibilities that could have also been potential candidates for the final outcome.

Things get even more interesting when one examines the impact an internalist versus an externalist view of autonomy has on an agent's history. The internalist view only sees an agent's history as relevant to his autonomy in so long as it "yield[s] rationality, an ability to acquaint oneself with relevant facts, reliable capacities for decision-making and action, current psychic integration, and the like" (Mele 2009, 4). Mele explains "[g]iven that the traits and capacities are in place and are exercised with appropriate care and suitable frequency, all else is irrelevant to psychological autonomy, including how the agents came to be as they are" (2009, 4). Dworkin continues his view on autonomy mentioned above - "[A]utonomy is a second-order capacity to reflect critically upon one's first-order preferences and desires, and the ability either to identify with these or to change them in light of higher-order preferences and values" (Dworkin 1988, as cited in Mele 2009, 3) by explaining that "[b]y exercising such a capacity we define our nature, give meaning and coherence to our lives, *and take responsibility for the kind of person we are*" (Dworkin 1988, as cited in Mele 2009, 5, my italics). Internalism views the cause of one's capacities and abilities as an independent issue from whether one performs an

4. This is Peter van Inwagen's (1983) definition of determinism, as cited in Mele 2009, 3.

act freely and thus is morally responsible for the action (Mele 2009, 5). This separation of cause and effect could be promising, as it would allow the compatibilist to assert that while the initial or background causes (one's history) are deterministic, free will lies in the intermediate step between the cause and effect. This is similar to Libet's veto power but with a longer interval than a few milliseconds to make changes. This intermediate step is when one critically reflects on one's first-order preferences and desires and decides whether to act in accordance with these desires, or to pursue a different course of action. Acting in accordance with the desire or pursuing a different action would be the effect, and it does not directly flow from the deterministic causal forces. It is one's second-order choice, and thus, he is responsible for his choice as he made it autonomously.

However, internalism is not without its critics, for the externalist view of autonomy is very interested in how an agent came to be. In this view autonomy depends on the agent's causal history, such as how she came to possess the desires and values that guide her self-reflection and decision making (Mele 2009, 4). The concern here becomes evident when two agents possess all of the (non-historical) qualities the internalist would require – reviewing facts, using reason in the decision-making process, critically examining their first-ordered desires, etc. – but their histories are so radically different that "we would be strongly inclined to regard one as significantly less autonomous than the other" (Mele 2009, 5). For example, say Patient A has a high level of external shame and frequently gets caught in a shame-pride cycle. The shame she feels from what she perceives as other peoples' negative perceptions of her leads her to restrict her food intake, which results in significant weight loss. This weight loss encourages her to continue restricting her food as she feels proud of her reduced body weight and control over her body. Patient B has difficulty accurately gauging her body size due to an assessment distortion in her self-perception, in addition to experiencing a decreased desire for food as a result of a dysregulation in the reward-processing mechanisms in her brain. Differences in histories such as these could help explain why some anorexics are able to get better and others are not, because some anorexic's histories (the deterministic events that lead to her illness) are too much for her to overcome – the deterministic deck is stacked against her, so to speak. When the factors that comprise one's causal history make free will seem implausible, is it possible to retain an internalist notion of free will that is more than just wishful thinking?

3. A Solution

Dworkin and Mele argue that in order to have autonomy we do not have to be completely independent of deterministic forces and I agree. There may be a way to reconcile the internalist and externalist views of autonomy. We can argue, like Mele, that if determinism is true and it is compatible with personal autonomy and free will we may not be able to take responsibility for our character but one can still have autonomy by "living in accordance with preferences and desires that one identifies with "in light of higher order preferences and values"" (Mele 2009, 6). We do not assume responsibility for our character as it is the product of external and deterministic causes over which we have no control and thus, no responsibility (Mele 2009, 6). However, these deterministic causes, our histories, need not undermine compatibilism for we still retain power (and responsibility) for our intentional actions. We may not choose what we've got, but we do choose what we do with it.

How does one exert control over her illness? A major step would be to seek treatment. Anorexia continues to have a poor prognosis (Galsworthy-Francis & Allan 2014, 55) and as such, the statistics for anorexics are grim. Steinhausen (2002) reports "[l]ongitudinal research has suggested fewer than 50% of individuals diagnosed with AN [anorexia nervosa] recover fully; 20–30% continue to experience residual symptoms, 10–20% remain significantly ill and 5–10% die from their illness" (as cited in Galsworthy-Francis & Allan 2014, 55). Morris (2008) states the mortality rates in anorexia are ten times that of the general population, and these are "the highest of all psychiatric disorders" (Harris & Barraclough 1998, as cited in Galsworthy-Francis & Allan 2014, 55). These statistics could be viewed as an agent's lack of free will or control over the illness; however it is interesting to note that individuals with anorexia tend to resist treatment and participation in treatment studies (Agras 2010, 488). Those who do seek treatment have a high rate of premature termination, the literature documents that rates of 50% are not uncommon (Sly et al. 2014, 40). It is beyond the scope of this paper to explore why premature termination rates are so high, as what is relevant and important for the task at hand is to demonstrate the anorexic *is able to make a choice*. If she can choose, opt, or decide whether or not to participate in treatment for her illness then she has free will as she has the ability to choose otherwise or contra to her first-order desires. The patients that prematurely terminate treatment typically do so at lower weights than those who complete treatment, "which is an indicator for the need for subsequent rapid readmission to [a] hospital" (Sly et al. 2014, 40). What is important to note is that these

patients *chose* to leave or not comply with treatment.[5] This choice defeats the second condition in the second premise of the Alternative Possibilities Condition:

1. The existence of alternative possibilities (or the agent's power to do otherwise) is a necessary condition for acting freely (of one's free will).
2. ~~Determinism is not compatible with alternative possibilities (it precludes the power to do otherwise).~~

About half of anorexic patients do get better or at least complete treatment – and half do not. The deterministic cause of one's illness does not preclude her from choosing from alternative possibilities - either to get better, or worse, or stay the same. This brings us back to Mele's assertion about physically possible futures. While only one of these outcomes is logically possible, there is nothing that necessitates which of these three outcomes will be determined because the choice is ultimately the anorexic's. She possesses the capacity to reflect critically on her first-order preferences and desires (such as to refuse food) and either identify with these desires or to change them (eat enough food so she is able to gain weight) via her higher-order desires. This is Dworkin and Mele's definition of autonomy and the anorexic fits the criteria.

This may actually be closer to Harris' argument than one would have initially thought. He argues:

> Becoming sensitive to the background causes of one's thoughts and feelings can – paradoxically – allow for greater creative control over one's life. It is one thing to bicker with your wife because you are in a bad mood; it is another to realize that your mood and behavior have been caused by low blood sugar. This understanding reveals you to be a biochemical puppet, of course, but it also allows you to grab hold of one of your strings: A bite of food may be all that your personality requires. Getting behind our conscious thoughts and feelings can allow us to steer a more intelligent course through our lives (while knowing, of course, that we are ultimately being steered). (2012, 47)

5. In some cases the premature termination is a staff initiated discharge which happens in cases where "the clinical team feel the patient is not engaging with the ethos of treatment, or not working in alliance with the boundaries of the treatment program. For example, a patient may disengage from the program and deliberately cease gaining weight" (Sly et al. 2014, 40). While the termination may be the decision of the clinical team, the choices that lead to this decision are the patient's.

This is along the lines of reconciling internalism and externalism, with the exception of one main difference – responsibility. I am willing to grant that the power of being able to "grab one of your strings" also gives you the ability to warrant responsibility for your actions. As such, we are able to remain in Smilansky's Community of Responsibility after all, but with a better justification than merely entertaining false beliefs. This, in turn, also permits the anorexic responsibility over her deterministic illness. She certainly did not choose or have any power over the underlying mechanisms that caused her eating disorder, but because she can choose her intentional actions she can be responsible for them. This of course seems more agreeable (and less problematic) when she is able to complete treatment and overcome her illness than if she is unsuccessful. But if we are able to attribute merit to an agent when she is successful, we must also attribute blame where it is warranted. The anorexic's autonomy and control can allow her to beat her illness. And indeed, some do.

References

Agras, Walter Stewart. 2010. "Chapter 27: Overview." In *The Oxford Handbook of Eating Disorders*, edited by W. Stewart Agras, 486–490. New York: Oxford University Press.

Alonso-Alonso, Miguel. 2013. "Brain, Reward, and Eating Disorders: A Matter of Taste?" *American Journal of Psychiatry* 170 (10): 1082–1085.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* Fifth Edition. Washington: American Psychiatric Publishing.

Benninghoven, Dieter, Lena Raykowski, Svenja Solzbacher, Sebastian Kunzendorf and Gunter Jantschek, 2007. "Body Images of Patients with Anorexia Nervosa, Bulimia Nervosa and Female Control Subjects: A Comparison with Male Ideals of Female Attractiveness." *Body Image* 4 (1): 51–59.

Cash, Thomas F. and Edwin A. Deagle, 1997. "The Nature and Extent of Body-Image Disturbances in Anorexia Nervosa and Bulimia Nervosa: A Meta Analysis." *International Journal of Eating Disorders* 22 (2): 107–125.

Galsworthy-Francis, Lisa and Steven Allan. 2014. "Cognitive Behavioural Therapy for Anorexia Nervosa: A Systematic Review." *Clinical Psychology Review* 34 (1): 54–72.

Harris, Sam. 2012. *Free Will*. New York: Free Press.

Kane, Robert. 2009. "Introduction: The Contours of Contemporary Free Will Debates." *The Oxford Handbook of Free Will Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195178548.001.0001/oxfordhb-9780195178548-e-1.

Libet, Benjamin. 2009. "Do We Have Free Will?" *The Oxford Handbook of Free Will Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195178548.001.0001/oxfordhb-9780195178548-e-25.

Mele, Alfred R. 2009. "Autonomy, Self-Control, and Weakness of Will." *The Oxford Handbook of Free Will Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195178548.001.0001/oxfordhb-9780195178548-e-24.

Morh, H.M., J. Zimmermann, C. Röder, C. Lenz, G. Overbeck and R. Grabhorn. 2009. "Separating Two Components of Body Image in Anorexia Nervosa Using fMRI." *Psychological Medicine* 40 (9): 1519–1529.

Roskies, Adina. 2006. "Neuroscientific Challenges to Free Will and Responsibility." *Trends in Cognitive Sciences* 10 (9): 419–423.

Sly, Richard, Victoria A. Mountford, John F. Morgan and J. Hubert Lacey. 201). "Premature Termination of Treatment for Anorexia Nervosa: Differences Between Patient-Initiated and Staff-Initiated Discharge." *International Journal of Eating Disorders* 47 (1): 40–46.

Smilansky, Saul. 2009. "Free Will, Fundamental Dualism, and the Centrality of Illusion." *The Oxford Handbook of Free Will Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195178548.001.0001/oxfordhb-9780195178548-e-22.

Troop, Nicholas A and Chloe Redshaw. 2012. "General Shame and Bodily Shame in Eating Disorders: A 2.5-Year Longitudinal Study." *European Eating Disorders Review* 20 (5): 373–378.

Troop, Nicholas A., Steven Allan, Lucy Serpell and Janet L. Treasure. 2008. "Shame in Women with a History of Eating Disorders." *European Eating Disorders Review* 16 (6): 480–488.