

2012

Autonomy and Paternalism

Ilona K. Phipps-Morgan
Scripps College

Recommended Citation

Phipps-Morgan, Ilona K., "Autonomy and Paternalism" (2012). *Scripps Senior Theses*. Paper 58.
http://scholarship.claremont.edu/scripps_theses/58

This Open Access Senior Thesis is brought to you for free and open access by the Scripps Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Scripps Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

AUTONOMY AND PATERNALISM

by

ILONA K. PHIPPS-MORGAN

**SUBMITTED TO SCRIPPS COLLEGE IN PARTIAL FULFILLMENT
OF THE DEGREE OF BACHELOR OF ARTS**

PROFESSOR WEINBERG

PROFESSOR TANNENBAUM

APRIL 20, 2012

Introduction

According to John Stuart Mill, “Neither one person, nor any number of persons, is warranted in saying to another human creature of ripe years, that he shall not do with his life for his own benefit what he chooses to do with it” (qtd. Dworkin, “Paternalism” 107). Thus Mill champions autonomy above all else in self-regarding cases. And indeed, we have come to treat autonomy as one of our highest values. The value that we confer to autonomy may be traced to various reasons, among them arguments that respecting others’ autonomous choices is the best way of securing their welfare, that respecting autonomy creates an environment that promotes individuality and human flourishing (Arneson 481), and lastly, that autonomy is simply part of what it means to be a person, and to deny others’ autonomy is to deny them some aspect of personhood (Dworkin, “Paternalism” 117).

And yet, autonomy is not our only value. In particular, I will consider the competing value of human welfare and the justifiability of overriding autonomy through paternalistic interferences. For example, are we justified in forcing a life-saving blood transfusion onto an unwilling Jehovah’s Witness? On the one hand, the Jehovah’s Witness may be exercising her autonomy in refusing to accept the blood transfusion, and we typically wish to respect others’ autonomy. Yet at the same time, we may want to override the Jehovah’s Witness’s autonomous decision on the grounds that her refusal — resulting in death in this case — is bad for her.

In order to lay the groundwork for discussing paternalistic interferences with autonomous decisions, I first consider different conceptions of autonomy, welfare, and paternalism, and determine which I mean to use. In particular, I proceed with Dworkin’s characterization of autonomy as a combination of authenticity and self-determination; Nussbaum’s capabilities theory in order to determine welfare; and a definition of paternalism as being an interference

with a subject's liberty or autonomy that is motivated exclusively by consideration for that subject's own good or welfare ("subject" will refer to the potential subject of paternalism).

Once I have working definitions for autonomy, welfare, and paternalism, I consider arguments justifying paternalistic interferences. Because I especially wish to determine when paternalistic interferences that conflict with a subject's autonomous decision are justified, I begin with Scoccia's arguments for using hypothetical consent — which is based on what would maximize the subject's welfare — to justify paternalistic acts. Using Scoccia's argument, I consider a few cases in which concerns for welfare may justify paternalistic acts overriding the subject's autonomy.

However, hypothetical consent does not go very far in justifying paternalistic acts. Therefore, I also consider arguments justifying paternalism in cases where the subject is not necessarily fully autonomous when making or acting upon a decision. For example, Carter argues that paternalistic acts are justified if autonomy has been waived through prior or subsequent consent. Additionally, I look at justifying paternalism when the subject's autonomy is compromised through involuntariness or incompetence.

What is autonomy?

To determine if or when autonomy can be overridden by other concerns, we must first define the conception of autonomy. The word "autonomy" literally means "the having or making of one's own laws," as derived from the Greek words for "self" and "law" (Feinberg, *Harm...* 27). However, this root word has come to refer to multiple different conceptions of autonomy. Feinberg groups these conceptions into four broad types. First, "autonomy" refers to the capacity to govern oneself; second, to the actual condition of self-government and its

associated virtues; third, to an ideal complex of character traits related to the condition of self-government; and finally, to the sovereign authority to govern oneself (*Harm...* 28). Of these four categories, I will consider the first two. I won't consider autonomy as an ideal, because here I am more concerned with autonomy as exercised in the real world — I wish to determine whether actual, rather than ideal, autonomy can be overridden by concerns for a subject's welfare. I also won't consider autonomy as a right, because simply dealing with the right to autonomy doesn't seem to provide sufficient background as to what it is about autonomy that makes it worth preserving and worthy of being given the status of "right" in the first place. Without such background, it will be difficult to measure autonomy against other values such as welfare.

The conception of autonomy as a capacity, on the other hand, doesn't reveal the value of autonomy per se, but is nevertheless helpful because one must have the capacity to exercise autonomy before it can be valued. This capacity is determined by one's ability to make rational choices (Feinberg, *Harm...* 28). Thus Christman writes that the basic capacity of autonomy involves the ability "to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces" (Christman¹). For reasons and motives to be one's own, they must come from an internal process of reflection and evaluation as opposed to simply being imposed by external forces such as manipulation or pressure (Christman). This definition of basic autonomy does not require that decisions come entirely from one's internal processes. It is meant to apply to most adults despite the fact that they may not be ideally autonomous (Christman) while admitting that children are not yet fully

¹ Christman's work is in a webpage, and so page numbers are not available.

autonomous beings, although they usually gain more autonomy as they grow and learn². Similarly, the criterion of being able to make rational choices, or being “competent,” tends to exclude the insane, severely retarded, the senile, and the comatose as well (Feinberg, *Harm...* 28). Although competence can be viewed as relative, such that normal adults may appear “incompetent” relative to the gifted, the competence required to possess the capacity of autonomy is typically measured relative to a threshold, such as a minimal level of intelligence or relevant capability with regard to a decision or task. Thus, a subject is competent if she satisfies this threshold (and thus she possesses the capacity for autonomy) or she’s incompetent if she fails to satisfy the threshold (in which case she doesn’t possess the capacity for autonomy) (Feinberg, *Harm...* 29).

Measuring competence by a threshold is important because it means that the gifted cannot simply interfere with normal people’s choices on the grounds that they are “incompetent.” Rather, if a subject meets the threshold for autonomy, then the interference of a more gifted person does constitute undermining the subject’s autonomy. That said, once the threshold is met, people’s capacities may differ in degree, and so some people who satisfy the threshold may be *less* competent even if they aren’t *incompetent* (Feinberg, *Harm...* 30). As a result, Feinberg notes, “the person whose relevant capacities are just above the bare threshold of competence that qualifies him for *de jure* self-government may rightfully rule himself, but in fact he may rule himself badly, unwisely, only partially” (*Harm...* 30). Consequently, even though interference with a competent subject’s choices may undermine her autonomy, such interference could nevertheless make her better off.

² In contrast, the *ideal* of autonomy may imply that a person is “maximally authentic and free of manipulative, self-distorting influences” — a condition that serves as a goal rather than a description of most adults (Christman)

The conception of autonomy as a capacity describes our actual ability to be autonomous, but it doesn't necessarily treat autonomy as having special value. Even though Feinberg implies that the capacity for autonomy somehow leads to a right of autonomy in claiming, as quoted above, that a person with the capacity for autonomy may "rightfully rule himself," he doesn't show why the capacity for autonomy has value in itself. After all, the simple ability to do something doesn't automatically make it valuable; e.g., I can enter a stranger's home by breaking through a window, but we don't value my ability to break into homes. Thus, it's not just having a capacity that's important. In some way that differs from my capacity to break into houses, the capacity for autonomy is especially worthy of preservation. Consequently, considering only the capacity for autonomy isn't sufficient for our purposes, as it doesn't seem to explain what it is about this capacity that we value.

The value of autonomy is partially illuminated by Feinberg's second characterization of autonomy as a condition. The condition of autonomy incorporates a variety of virtues drawn from a conception of self-determination (Feinberg, *Harm...* 32). These virtues include self-possession, distinct self-identity (individuality), authenticity, self-creation (self-determination), self-legislation, moral authenticity, moral independence, integrity, self-control (self-discipline), self-reliance, initiative, and responsibility for self (Feinberg, *Harm...* 32–43). In particular, I wish to focus on the virtues of authenticity and self-determination.

According to Feinberg, "a person is authentic to the extent that... he can and does subject his opinions and tastes to rational scrutiny... and can and does alter his convictions for reasons of his own" (*Harm...* 33). These opinions and tastes may be partly influenced by others, but the authentic person should evaluate and alter her opinions and tastes based on her own character as well. In contrast, forming her opinions and tastes merely based on another's (or society's) lead

would not be authentic and, on this account, the subject would fail to form her opinions and tastes autonomously (Feinberg, *Harm...* 32). Authenticity ties very closely to the condition of self-determination, under which a subject determines how her life goes according to her character and judgments (Feinberg, *Harm...* 33), as both of these require that the subject reflect rationally.

With respect to both of these conditions, Feinberg argues that we should not define authenticity and self-determination so strictly that they are impossible to satisfy. Feinberg acknowledges that “a person must already possess at least a rudimentary character before he can hope to *choose* a new one” (*Harm...* 33). This rudimentary character may be partially formed through one’s innate character, if such a character exists, but it is also inevitably the product of our upbringing. To require that one’s character be solely formed from innate traits in order to be authentic and self-determining would disqualify actual people from being authentic and self-determining. Thus, Feinberg suggests that we take a more flexible view under which we may judge a person to be authentic and self-determining — and, under these conditions, autonomous — even while recognizing that this person’s character has, unavoidably, been influenced by outside forces, especially during the early formation of the person’s character (Feinberg, *Harm...* 33).

For example, a newborn’s character is very undeveloped, and so the extent to which the newborn shapes her character is minimal (Feinberg, *Harm...* 34). Rather, at this point in her life, the newborn’s character is influenced more by external forces, although Feinberg does believe that the newborn’s own genes and early environment give the process of her character-development “its own distinctive slant” (*Harm...* 34). If parents are concerned for the autonomy of the adult that the newborn will grow into, then as this “slant” becomes more apparent, the

parents should respect that initial bias (Feinberg, *Harm...* 34). Thus, the child's character-development is a combination of outside forces and her own input. As she grows older and her character becomes more developed, the child's input will become more and more influential in her development "until at the arbitrarily fixed point of full maturity, [she] is at last fully in charge of [herself], [her] more or less finished character the product of a complicated interaction of external influences and ever-increasing contributions from [her] own earlier self" (Feinberg, *Harm...* 34). In this case, Feinberg considers the resulting adult to be self-determining despite the fact that her earlier development was necessarily influenced by external forces. Moreover, the resulting adult here is what Feinberg sees as the ideal of a "self-made person" (*Harm...* 35), which I take to be the closest approximation that could in reality be achieved. However, because I am concerned not with ideal autonomy, but with actual autonomy, it seems that a subject may be self-determining even if her parents do not raise her "with maximal regard" for her autonomy. She may be self-determining — perhaps not ideally so, but generally so — if she contributes to the development of her character as she matures and is, as an adult, able to evaluate and choose her character without being driven merely by outside forces. What is important for autonomy is not that the subject has never been externally influenced, but that she reflect rationally when considering her principles and the adoption of other candidate principles and that she have a role in the development of her character.

Dworkin echoes a very similar view of autonomy when he writes, "The idea of autonomy is not merely an evaluative or reflective notion, but includes as well some ability both to alter one's preferences and to make them effective in one's actions and, indeed, to make them effective because one has reflected upon them and adopted them as one's own" (*The Theory...* 17). In essence, Dworkin is adopting a view of autonomy based on authenticity and self-

determination in calling for evaluating and reflecting upon one's preferences or judgments. By reflecting upon and adopting her preferences as her own, the subject is authentic based on Feinberg's description of authenticity as the condition of subjecting her opinions and tastes to rational scrutiny and altering them accordingly. Similarly, in making her preferences effective in her actions, the subject is self-determining because she is involved in creating her character and determining the course of her life.

However, Dworkin elaborates further and suggests that, in order to satisfactorily evaluate and reflect upon her preferences, the subject must consult her second- or higher-order desires (*The Theory...* 15). This is because if we only consider first-order desires, we ignore "a crucial feature of persons, their ability to reflect upon and adopt attitudes toward their first-order desires, wishes, intentions" (Dworkin, *The Theory...* 15). For example, Dworkin considers the case of Odysseus. In his travels, Odysseus approaches a sea known to harbor sirens that sing so beautifully that sailors are driven to dive off their ships, only to drown in the waters. Desiring to hear the sirens' song but not to lose his life as a result, Odysseus commands his men to bind him to the mast of the ship and to ignore his orders to be untied while he's under the sirens' influence. In this example, Odysseus' commands to be untied are involuntary³, but aside from consideration of voluntariness, Dworkin claims that "there is another dimension of [Odysseus'] conduct that must be understood," namely that Odysseus not only has first-order preferences to be untied and to pursue the sirens, but he also has preferences about these preferences (*The Theory...* 15). By ignoring the commands reflecting his first-order preferences, Odysseus' men actually promote Odysseus' ability to be authentic and self-determining.

³ Voluntariness will be considered in more detail later, but in the story of Odysseus, it is assumed that the sirens' song temporarily drives men to act in ways not according to their own judgments and so they aren't voluntary.

Dworkin takes Odysseus' ability to evaluate and to identify with or reject his lower-order preferences and reasons for which he acts as crucial to his authenticity (*The Theory...* 15).

Although a subject may not be able to change her first-order preferences at will, to be authentic she should be able to subject them to rational scrutiny and to change them if there is sufficient reason (Dworkin, *The Theory...* 16), where "sufficient reason" is presumably defined under circumstances of rational scrutiny. Thus, authenticity is not based simply on the subject's character and desires, but on whether the subject is capable of deliberating on the reasons for which she acts, requiring the use of higher-order desires (Dworkin, *The Theory...* 15).

It's important to recognize that in evaluating her lower-order desires, the subject is supposed to exercise some rationality — Dworkin calls for "critical" reflection upon first-order preferences, desires, and wishes (*The Theory...* 20) — but not necessarily *perfect* rationality. In contrast, Feinberg characterizes Rawls as emphasizing one's rational will such that one's autonomy depends not on one's *actual* reflection and consent, but on one's hypothetical and rational consent (Feinberg, *Harm...* 36). "Thus," Rawls states, "acting autonomously is acting from principles that we would consent to as free and equal rational beings, and that we are to understand in this way" (qtd. Feinberg, *Harm...* 36). According to this view, Feinberg explains that a person acts autonomously only if her actions are what hypothetical persons acting rationally would do; and if the person were herself rational, she would act the same way (*Harm...* 36). "No matter that a person does not in fact consent to the rational principles," Feinberg writes. "What is required is that hypothetical persons in certain circumstances *would* consent to them, and presumably he would too if only he were more rational" (Feinberg, *Harm...* 36). Because it is not the subject's actual process of evaluation and decision-making that counts but simply *what* she would choose if she were fully rational, it seems that according to this view,

autonomy is measured more by the content of a decision or action than by how the subject arrives at the decision or action. Thus, Feinberg claims that on Rawls' view, persons cannot be simultaneously autonomous *and* wicked, dishonorable, or selfish, for hypothetical, rational persons would not act wickedly, dishonorably, or selfishly (*Harm...* 36).

I reject the view of autonomy as being measured by the choices of hypothetical, rational persons for similar reasons that I rejected the view of autonomy as an ideal: I am more concerned with autonomy as exercised by actual people in the real world. For instance, suppose a subject critically reflects upon her preferences and judgments and, concluding that these judgments and preferences are in line with her character, she proceeds to act according to her judgments and preferences; that is, the subject acts authentically and is self-determining. However, her action fails to correspond to what a fully rational person would do. In this case, I believe that interference with the subject's action disrespects her autonomy, whereas presumably under Rawls' view, the subject's action isn't truly autonomous in the first place because it doesn't correspond to what a fully rational person would do. Requiring that a subject's action align with that of a fully rational person seems to strip many, if not most or all, subjects of their autonomy. Here I desire a conception of autonomy that is applicable to actual people, even if the content of their choices doesn't appear ideal, and so I reject Rawls' conception.

Instead, I will adopt Dworkin's view that autonomy consists of both authenticity and self-determination, where authenticity refers to the reflection upon and evaluation of one's preferences and judgments as well as the ability to alter these preferences based on one's evaluations, while self-determination refers to being able to determine one's life in accordance with one's character and judgments. I find Dworkin's view attractive because he explains *how* autonomy is exercised — by reflecting upon and evaluating one's desires and acting accordingly

— and so it is possible, to some extent, to differentiate cases in which a subject is autonomous from one in which she is not autonomous.

For example, if I reflect upon my values and my goals for myself and, in accordance with these values and goals, I choose to become an entrepreneur, then my pursuit of being an entrepreneur is autonomous. In choosing my career, I act authentically by reflecting upon and evaluating my values, while I am self-determining by acting upon my considerations in the pursuit of this career. Because I am both authentic and self-determining, I am autonomous. On the other hand, perhaps I have reflected upon my values and believe that becoming an entrepreneur is most in accordance with my character, but I nevertheless end up going to medical school because I am forced to do so or perhaps because I feel that I owe going to medical school to my parents (if, for instance, they have given me an excellent education and expect me to become a doctor, and I feel obliged to make them proud). In this case, although I may be authentic, I am not self-determining because, forced by my obligations, I do not determine my life in accordance with my considered judgments. Similarly, if my parents pressure me into valuing membership in the medical profession and I then choose to follow their advice, I may be self-determining, but my values aren't authentic because they aren't the product of my own critical reflection and evaluation. Of course, it is possible that my parents are so convincing that my higher-order values support the change in my value for being part of the medical profession — after all, it is still possible for someone to authentically change her mind in response to other's arguments, provided that her change is ultimately caused by her own reflection and evaluation of values.

In addition to explaining how autonomy is exercised, Dworkin claims that autonomy is valuable because by evaluating and forming authentic judgments and then using them to

determine one's actions, "persons define their nature, give meaning and coherence to their lives, and take responsibility for the kind of person they are" (*The Theory...* 20). That is, autonomy is simply part of what it means to be a person. John Stuart Mill adopted a similar view; as Dworkin writes, "It is because coercing a person for his own good denies [him] status as an independent entity that Mill objects to it so strongly and in such absolute terms. ... It is the privilege and proper condition of a human being, arrived at the maturity of his faculties, to use and interpret experience in his own way" (Dworkin, "Paternalism" 117). Thus, we value autonomy because it is part of who we are and how we see ourselves as individuals and persons. Furthermore, Mill claimed that respecting autonomy is vital in creating an environment that promotes individuality (Arneson 481). By respecting each other's and our own autonomy, we create an environment conducive to developing our own, unique life plans (Arneson 481).

There are several alternatives to an autonomy-based moral theory, such as theories based on central values of care, happiness, or virtue (Christman). Here, I specifically wish to consider the competing value of welfare. Many, if not most, philosophers — including both Mill and Kant — recognize the value of welfare, although they often place the value of autonomy more highly. But do they value autonomy *too* highly? Herzog points out that exclusive consideration for autonomy may seem callous (151). The policy of neutrality in a liberal state may imply that "We know what the good is, or we know at least some things it isn't, but we're going to respect your rights and let you make the wrong choice" — this "hardly seems an inspiring public ethos," Herzog muses (151). While we may welcome the freedom afforded by state neutrality, the silence of the neutral state may also be seen as "loathsome, a public declaration of our yellow-

bellied relativism, or a cruel way of condemning individuals to spend their lives groping uncertainly for leadership that the state should provide” (Herzog 148).

Given widespread (although not universal) opposition to laws that attempt to regulate choices that many believe should be individualized, such as those regarding diet, we often seem to accept the flaws of a neutral state in exchange for the freedom it grants us. Yet many people simultaneously embrace some paternalistic interferences with our autonomy, such as seatbelt laws or the prohibition of highly addictive and dangerous drugs. These are cases in which our welfare may seem to override our right to autonomy. Admittedly, one could consider seatbelt laws, for example, as simply enforcing an action (wearing seatbelts) that we would choose for ourselves but are too weak-willed to consistently do. However, I don’t believe that all paternalistic laws can be conceived of in this matter; specifically, it seems that the prohibition of highly dangerous drugs is enforced not merely because we personally would otherwise be too weak-willed to resist drugs, nor solely to protect us from the ill effects of others’ drug addictions, but also in part because we care for others’ welfare and wish to intervene on their behalf as well. In order to determine if and when concern for a subject’s welfare can justifiably outweigh her right to autonomy, we must first understand what is meant by “welfare.”

Welfare

In the appendix to *Reasons and Persons*, Derek Parfit considers three main kinds of theories about self-interest or what is best for someone (which I take to be her “welfare”): Hedonistic Theories, Desire-Fulfillment Theories, and Objective List Theories (493). According to Hedonistic Theories, that which would make the subject happiest, i.e. have the most pleasant mental state, is best for her (Parfit 493). In what Parfit considers the most compelling Hedonistic

Theory, Preference-Hedonism, a subject's well-being is considered in terms of her desired mental states (Parfit 493). Maximizing a subject's desired mental states and minimizing her undesired mental states increases her welfare. Thus, the more desired a mental state is, the more the subject's welfare is increased; the more undesired a mental state is, the more the subject's welfare is decreased. In order to maximize a subject's welfare, then, one should aim to fulfill her strongest desire, whether that be past, present, or future. So, for example, when Freud refused anesthetics near the end of his life because he preferred thinking in torment rather than being "confusedly euphoric," not using anesthetics was better for him (Parfit 494). If one adopts Preference-Hedonism as one's theory of welfare, interfering with the fulfillment of a subject's desires would, by definition, be worse for her.

However, in some cases it is possible to interfere with a subject's choices in a way that does help fulfill her desires and therefore is good for her. First of all, a subject's past or future desires may be stronger than her present ones. For example, although a high school student may presently desire to skip class and not do her homework, her future desire to go to a good college may be stronger than her present desires. In this case, assuming that attending class helps to fulfill her future desire of going to a good college, interfering with her choice to skip class increases her welfare.

Secondly, in assessing the subject's desires, the Preference-Hedonist includes hypothetical in addition to actual desires (Parfit 496). Hypothetical desires refer to those desires that a subject would have if she had chosen otherwise. For example, Parfit considers a subject with two choices: staying home and going to a party. Perhaps the subject chooses to stay home and read, and her present desire to have a peaceful evening is thereby fulfilled. Yet the subject also had the choice of going to a party, and so one can consider the counterfactual case in which

she does choose to go to the party. Under these circumstances, the subject would desire to have an exciting and fun evening, and she does. The desire to have a fun evening that she *would* have if she had chosen to go to the party is her hypothetical desire.

Either course of action — staying home or going to the party — would result in a pleasant evening for the subject that she would not later regret (Parfit 496). Because the subject's desires are thereby fulfilled, her welfare is increased. However, if going to the party would fulfill the subject's desires only a little, whereas staying home would fulfill her desires more, then Parfit argues that we should still be able to claim that staying home is the better alternative with regard to the subject's welfare (496). Thus, in Preference-Hedonism, in order to maximize one's welfare, one should fulfill the desire — whether it be actual or hypothetical — that is strongest (Parfit 496). Furthermore, if the subject has more available alternatives at the time (perhaps instead of staying home or going to the party, she could also choose to visit her sister or to walk her dog), then one must also consider the hypothetical desires that the subject would have in any of these available alternatives (Parfit 496). If there is more than one “strongest” desire, then choosing to satisfy any of those strongest desires is appropriate for maximizing welfare.

The second set of theories considered by Parfit is Desire-Fulfillment Theories, which claim that “what would be best for someone is what, throughout his life, would best fulfill his desires” (493). Again, these desires may be past, present, future, or hypothetical. There are two main Desire-Fulfillment Theories. According to the first, the Unrestricted Theory, we consider *all* of the subject's desires, irrespective of whether they are about her life or others', in order to determine what is best for the subject (Parfit 494). Parfit rejects this theory and claims that considering those desires that don't actually pertain to one's life is implausible. For example, if I

meet a stranger who has a terrible disease, my sympathy may be aroused at the time, and I may desire that this stranger be cured. However, if it later so happens that the stranger is cured, then, Parfit argues — contra the Unrestricted Theory — *my* life hasn't actually improved by the satisfaction of this desire because the desire isn't about my own life (494).

In some sense, it might seem that my desire that the stranger be cured *is* about my life; one could say that I wish to lead a life in which the stranger is cured. To illustrate the difference between this type of desire and desires that are “actually” about my life, Parfit likens the former desires to so-called Cambridge-changes. Cambridge-changes are changes in true statements that can be made about an object but which are opposed to “real” changes in the object itself. For example, Parfit invites us to “suppose that I cut my cheek while shaving. This causes a real change in me. It also causes a change in Confucius. It becomes true, of Confucius, that he lived on a planet in which later one more cheek was cut. This is merely a Cambridge-change” (494). If I cut my cheek while shaving, there is a change in the true statements that can be made about Confucius, but Confucius wasn't changed in and of himself. Similarly, if a stranger is cured, then I am not changed, and therefore my life isn't made better by the stranger's recovery. In contrast, desires about my own life typically involve something I wish to have or do or the kind of person I wish to be⁴. Thus, because the stranger's recovery does not affect something I wish to have or do or the kind of person I wish to be, my life isn't made better by his recovery. It is not important that the cured person be a stranger to whom I have no other ties; if a dear friend happens to recover from a disease, *my* life is still not made better by his recovery. However, if I invest myself in helping a friend recover from a disease, and she does subsequently recover, then my life is made better. Through the success of my efforts, my desires regarding the kind of

⁴ This way of distinguishing desires about my own life was stated by Julie Tannenbaum, who drew on ideas from Nozick.

person I wish to be (e.g. one who is there for her friends) are satisfied, thereby increasing my welfare.

Instead, Parfit finds Success Theory, another Desire-Fulfillment Theory, more compelling. In Success Theory, only the satisfaction of those desires about one's own life contribute to one's welfare (Parfit 494). This theory is similar to Preference-Hedonism except in that Success Theory "appeals to all of our preferences about our own lives" whereas Preference-Hedonism "appeals only to preferences about those present features of our lives that are introspectively discernible" (Parfit 494). To illustrate the distinction, Parfit considers the desire not to be deceived by others. According to Preference-Hedonism, actually being deceived doesn't make me worse off if I'm not aware of the deception; all that matters is my state of mind, not the reality of my situation (Parfit 494). In contrast, according to Success Theory, it is reality that determines how well off I am. Merely believing that others aren't deceiving me doesn't make me better off if I am in fact deceived (Parfit 494).

As with Preference-Hedonism, interfering with a subject's preferred decision generally won't increase the subject's welfare under Success Theory. However, again interference may increase a subject's welfare in cases where the subject's past, future, or hypothetical desires are stronger than her present ones. In addition, if the subject's decision is intended to satisfy one of her desires but, in fact, it won't (i.e. she's mistaken), then interference that fulfills her desires could also promote her welfare. In contrast, under Preference-Hedonism, if the subject falsely believes that X is true, and as a result she has her desired mental state, then her welfare is increased even if X is not, in fact, true. Consequently, under Preference-Hedonism, it would not necessarily promote a subject's welfare to disillusion her and thus remove her desired mental state. Indeed, in Preference-Hedonism, because the subject's mental state determines her

welfare, actions that trick her into believing that X is the case and thereby give her desired mental states could potentially make her “better” off. Because Success Theory doesn’t only consider the subject’s mental states (does the subject believe that she’s achieving what she desires to achieve?), but also the actual, external satisfaction of her desires (is she, in fact, achieving what she desires to achieve?), I find Success Theory the more attractive of the two.

However, both Preference-Hedonistic and Desire-Fulfillment Theories may be considered problematic for several reasons. First, insofar as one’s preferences or desires are shaped by outside forces such as society, culture, or even past actions that give rise to preferences or addictions, it seems questionable to claim that satisfying these desires can increase one’s welfare. For example, consider a society in which women have a lower social status and consequently men have greater opportunities than women. If women are raised to believe that such a social order is desirable, then according to Preference-Hedonism or Desire-Fulfillment Theories, being given fewer opportunities increases the women’s welfare. Yet it seems that promoting inequality cannot truly be in the women’s best interests.

A second worry, expressed by Cass Sunstein, is that many of our preferences are “endogenous,” meaning that they are malleable and may change depending on the context in which the preference is expressed, past consumption choices, and cultural and legal factors (5). For Sunstein, who wishes to consider the proper extent of government intervention with people’s choices, endogenous preferences are worrisome because if preferences aren’t fixed, then a democracy that wishes to maximize welfare gains cannot simply apply one set of rules for all time (10). Indeed, the very rules or policies created by the government may themselves influence people’s preferences. Thus, for example, giving women the right not to be sexually harassed will impact people’s attitudes towards sexual harassment (8). However, although the

changeability of desires makes rule-setting difficult, it is not itself an objection to calculating welfare based on desire fulfillment.

Another, more serious, challenge posed by the malleability of desires is the implication that, in order to increase a subject's welfare, one could either satisfy her desires *or* change her desires in order to align with her current situation. In some cases, if the subject's desires are overly optimistic (e.g. if the subject desires something extremely difficult and unlikely to attain such as learning to fly), then it may indeed be best for her to reevaluate her desires and adjust accordingly. Yet in other cases, changing a subject's desires to lower her standards does not intuitively seem to increase her welfare. Returning to the example of an unequal society, suppose that many women *are* currently dissatisfied with the inequality of opportunities available to men and women — that is, unlike the previous example, the women desire equality despite the inequalities with which they were raised. Women's welfare could be increased by eliminating the inequality, or women could be encouraged to change their desires and embrace the current state of affairs instead. If women embraced a lower social status, then their new desires would be satisfied, and accordingly, on Preference-Hedonistic or Desire-Fulfillment Theories, their welfare would seemingly be "increased." This conclusion is counter-intuitive in that, although the women's (altered) desires are now satisfied, their sub-optimal condition of living in an unequal society remains the same.

The third main objection to Preference-Hedonism and Desire-Fulfillment Theories is that people's desires may simply be bizarre, in which case fulfilling those desires doesn't necessarily seem to promote their welfare. For example, Parfit imagines a case where a man wants to spend his life counting blades of grass when he knows that he could instead achieve significant, beneficial results pursuing applied math (Parfit 500). On Preference-Hedonism or Desire-

Fulfillment Theories, if the man's strongest desire truly is to count blades of grass, then counting blades of grass would be better for him, a conclusion that may be difficult to accept.

On the one hand, I am not entirely convinced that pursuing math is indeed better for *this* man. Certainly it would be better for *others* if this man achieved beneficial results in math, and the results could benefit the man as well. But if the man truly desired to count blades of grass and found great satisfaction in doing so, then — assuming that he didn't later come to seriously regret all the time wasted on grass — I find it somewhat plausible that this man might be better off counting blades of grass. Again, this pursuit isn't in the best interests of others, but here I'm only concerned with the man's own welfare. If the satisfaction of his desire conflicted with other's interests (e.g. if the man's desire weren't to count blades of grass, which in and of itself is fairly harmless, but instead were to cause pain to others), then we may well be justified in interfering with the fulfillment of his desires in order to protect others. But where only the man is concerned, it may seem that if he wants to count grass, then perhaps he is best off doing so.

On the other hand, a man that chooses to count blades of grass for long periods sounds crazy, in which case rather than sitting back and accepting his desire to count grass, it would be more to the man's benefit if one helped him to reach a healthier state of mind. Yet Preference-Hedonism and Desire-Fulfillment Theories *do* admit that the man's welfare can be increased through intervention if the man has past, future, or hypothetical desires — in this case, his hypothetical desires could be drawn based on his alternative choice of pursuing math rather than counting blades of grass — that are stronger than his present ones. For example, if one intervenes to help the man reach a healthier mental state and his subsequent desires to reach mathematical greatness are stronger than his current ones to count grass, then it is better for the man to be thwarted in his current desires in favor of his future ones. If, however, the man's

present desire to count grass truly outweighs all his others — or perhaps he is permanently crazy and his past, future, and hypothetical desires all involve counting grass — then I am not completely convinced that interfering with his grass-counting in favor of pursuing math really is better.

Regardless of the third problem with Preference-Hedonism and Desire-Fulfillment Theories, the first two problems — first, that many preferences are shaped by culture and do not necessarily reflect the subject's best interests, and second, that changing a subject's desires to fit reality rather than changing reality to satisfy her preferences do not, contrary to these theories, seem to equivalently promote her welfare — stand. As a result, we need a theory that does not simply consider a subject's subjective values. Parfit's response is to consider Objective List Theories. According to these theories, some things have value for us irrespective of our desires. Some things simply are good for us and others aren't (Parfit 493). Examples of good things could be moral goodness, rational activity, the satisfaction of basic needs such as the need for nourishment or emotional connection to others, the development of one's abilities, knowledge, and awareness of true beauty, while examples of bad things could be betrayal, manipulation, slander, deception, the deprivation of liberty or dignity, and enjoying sadistic pleasure (Parfit 499). Although such good and bad things might align with good and bad things under Success Theory (e.g. in both cases, deception is bad), the value of things is determined in different ways. Specifically, Parfit writes that "On the Success Theory it is, for instance, bad for someone to be deceived if and because this is not what he wants. The Objective List Theorist makes the reverse claim. People want not to be deceived because this is bad for them" (499). Thus, the Objective List Theorist may respond to the man's desire to count grass by arguing that counting grass (at least as a long-term goal) is not objectively good for him because, drawing from Parfit's

examples of good things, counting grass is not rational activity, nor does it satisfy the man's basic needs or develop his abilities or knowledge.

The problem with Parfit's Objective List Theory is that he doesn't specify how these "objective" values are known, and what they draw their value from if not our own subjective values. Why, for example, is rational activity good for the grass-counting man if he doesn't value rational activity?⁵ Nussbaum answers this objection by claiming not that there are objective values independent of humans, but that there *are* some values that humans share — lending these values a sense of objectivity — which can be found by looking to history and human experience (69). As she writes, even if one claims that "we are always dealing with our own interpretations anyhow, [one] must acknowledge that universal conceptions of the human are prominent and pervasive among such interpretations" (Nussbaum 69). Rather than claiming a universal conception of the human "as part of the independent furniture of the world," Nussbaum makes the more modest empirical claim that, when one looks to history and human experience, one does find universal ideas of the human being, which can be used as a starting point for identifying common notions of the human good (69). For example, Nussbaum cites Ghanaian philosopher Kwame Anthony Appiah, who was raised in two cultures, visiting grandparents in Ghana and others in rural England among relatives in further countries. Through his interactions with people from numerous cultures, Kwame doesn't notice "unbridgeable alien 'otherness'" so much as "a great deal of human commonality" (69). Nussbaum admits that conceptions of humanity can and do differ between societies but nevertheless affirms that there is enough overlap in these conceptions to outline a notion of the human good (74).

⁵ Parfit ultimately embraces a view of welfare that is a composite of Desire-Fulfillment Theories and Objective List Theories. However, because he doesn't explain how objective values are known, a composite containing such values is also unsatisfying.

In particular, Nussbaum identifies that as humans, we share similar bodily constraints and needs such as hunger and thirst, need for shelter, sexual desires, and mobility; the capacity for pleasure and pain; cognitive capability, including perceiving, imagining, and thinking; early infant development; practical reason; affiliation with other humans; “relatedness” to other species and nature; humor and play; and separateness (75–79). Admittedly, these needs may not be shared by each individual human. However, typically those individuals who do not share these needs — e.g. a person who has no affiliation with other humans (perhaps an absolute hermit or a sociopath) — are seen as odd or crazy, outside of the typical societal norms. Thus, even if individuals do not all share these needs, as a whole, different societies do.

Moreover, because we do share these traits as humans, Nussbaum believes that we can further identify capabilities without which a life would be “too lacking, too impoverished, to be human at all” — and hence, neither could it be a good human life (80). As examples of capabilities needed for a good human life, Nussbaum offers, among others, the capabilities of being able to live the length of a normal human lifespan, being able to have good health, being able to avoid unnecessary and non-beneficial pain, being able to use the senses, and being able to form a conception of the good and to reflect critically on the planning of one’s own life (83–84). Because she treats this list as one of separate components, a good life is only attainable by ensuring each of the capabilities; that is, one cannot have a good life by trading off one capability for another (85–86). Having each of these capabilities is necessary to having a good life.

How, then, does this apply to the man whose deepest desire is to count grass? If counting grass endangers any of the capabilities necessary for a good life, then by definition, the man cannot have a good life counting grass. And indeed, simply based on the capabilities given above, choosing to spend his life counting grass quite possibly implies that the man does not

currently possess the ability to have good [mental?] health and the ability to form a conception of the good and to reflect critically on the planning of his life. The decision to count grass blades may not be, in and of itself, detrimental to the man's welfare, but it indicates that he doesn't currently possess the capabilities for a good life. Therefore, if one were considering the man's welfare, then rather than simply allowing the man to count grass as per Success Theory, one might first wish to ensure that the man possessed the requisite capabilities for a good life. Once he possesses these capabilities, however, Nussbaum allows that he can choose whether or not to exercise his capabilities. For instance, she points out that even though the need for sustenance is common to humans and the corresponding capability of having sufficient food is required for a good life, "a person with plenty of food can always choose to fast" (94). On her capabilities theory, the values concern what one is able to do; having these capabilities, one may then act according to one's subjective desires and preferences.

I find Nussbaum's formulation of objective values attractive because she provides for certain criteria for a good life which are objective in the sense that they are argued to apply to all humans. However, her support for these criteria is not based on a metaphysical nature of humans, but on empirical evidence from human experience indicating that across history, there are certain needs and values that people share. Using these "objective" criteria, we can escape entirely subjective valuations that might, for example, encourage women to accept inequalities for the sake of their "welfare." On the other hand, by citing capabilities rather than concrete actions as the criteria for a good life, Nussbaum acknowledges the importance of individuals' subjective desires with regard to their lives. Although certain capabilities are necessary for a good life, beyond the satisfaction of these capabilities, it is the fulfillment of the individuals' subjective desires that further increases or decreases their welfare. Thus, to me Nussbaum's

capabilities theory provides the best balance between considering the objective reality of the situation and respecting a subject's personal goals and wishes. Therefore, I will use Nussbaum's capabilities theory as the grounds for evaluating a subject's welfare. In addition, this theory opens the possibility of intervention on behalf of the subject's welfare in cases where either the subject doesn't currently possess the capabilities for a good life (as with the grass-obsessed man) or the subject's decision doesn't actually best reflect her subjective desires. Whether such interventions are actually justified in overriding the subject's autonomy remains to be seen.

Now that we have conceptions of autonomy and welfare, I will consider conflicts between these two values. Such a conflict may involve separate people's autonomy and welfare — for example, my autonomous desire to pickpocket someone conflicts with my victim's welfare. Alternatively, the conflict may involve the autonomy and welfare of a single person, which is the category of conflict I wish to consider. Thus, if I choose to drive without a seatbelt, this choice conflicts with my welfare if I end up in an accident and am injured because I wasn't wearing my seatbelt. In a liberal state, laws (e.g. some privacy laws) generally rule in favor of a subject's autonomy rather than her welfare; what a subject chooses to do to herself, as long as it doesn't interfere with others' welfare, is her own choice to make and act upon, even if the choice doesn't promote her welfare. Aside from formal laws, people in our society also frequently respect others' autonomous self-regarding choices that don't promote welfare. If someone wishes to eat at McDonald's every day, then we may allow her to act upon this choice even though we know that eating at McDonald's every day is unhealthy and likely won't increase her welfare (at least not in the long run).

We are frequently willing to defer to people's autonomous choices that seem to be to the detriment of their welfare. In part, this is because we, however well-meaning, do not know a subject's feeling and circumstances as deeply as does the subject herself, and so interfering is more likely than not to diminish the subject's welfare rather than increase it. Overall, we are all better off respecting each other's autonomy. Yet this argument fails to prohibit all cases of paternalism. Although a subject may usually be the best judge of her welfare, she nevertheless may make decisions that others can clearly see are contrary to her welfare, decisions that she will predictably regret later. Thus, both in laws and in our relationships, we may wish to interfere with people's autonomous decisions that are detrimental to their welfare. The wish to interfere with others' autonomy in favor of their welfare can be seen in such laws as those requiring the use of seatbelts or in cases where, for example, one friend may try to intervene in another's ill-fated relationship. Typically, the practice of interfering with someone's autonomy in favor of her welfare is considered paternalism. Eventually, I wish to consider when paternalism is justifiable, but first I will consider more specific accounts of what exactly should be included under the term "paternalism."

Paternalism

Typically, paternalism is roughly defined as an interference with a subject's liberty or autonomy that is motivated exclusively by consideration for that subject's own good or welfare (Arneson 471; Dworkin, "Paternalism" 108). This interference generally takes the form of coercion or manipulation, although Dworkin argues that "there are no methods of influencing people that are necessarily immune to being used paternalistically" (Dworkin, "Paternalism: Some..." 107). Shiffrin takes an even broader view of paternalism based not on the means or

even results of the interference, but on the interferer's motives. I'll consider the steps that Shiffrin takes to formulate her characterization of paternalism in order to better articulate what paternalism entails.

Shiffrin pursues her characterization of paternalism by assessing "what is central in our normative reactions to paternalism" and using this assessment to create a conception of paternalism that "complements and makes intelligible our sense of paternalism's normative significance" (212). That is, in order to characterize paternalism, Shiffrin considers cases of possible paternalistic action and then evaluates whether such cases arouse the reaction that we associate with standard cases of paternalism (such as coercing a subject for her own welfare) and whether these cases make sense of what it is about paternalism that we feel matters. By evaluating possible actions not on specific features (such as means or results) connected to them but on our normative reactions to the actions, Shiffrin delves into the significance of features typically associated with paternalism, such as freedom, the subject's will, and the paternalist's motivation.

To begin, she considers the connection of paternalism and freedom. Shiffrin admits that paternalism and freedom are strongly connected but claims that paternalism mustn't necessarily involve violating an autonomy right, interfering with a subject's liberty, or even simply diminishing the subject's freedom (213). Instead, paternalistic action may consist of an omission or refusal to do something (Shiffrin 213). For example, Shiffrin offers, suppose that I am planning to build a set of shelves and ask an acquaintance for help. Perhaps my acquaintance would be happy to help build a set of shelves, but she feels that I ask for help too much and am consequently not given the chance to learn necessary skills for myself. That is, she feels that I would actually be better off building the shelves myself, and on these grounds, she refuses to

help me. In this case, my autonomy and freedom haven't been affected — my acquaintance's refusal to help doesn't make me any less autonomous or free to do as I please than I was before I asked for her help — but Shiffrin argues that my acquaintance's refusal is nevertheless paternalistic because she is substituting her judgment for mine on what would be good for me (213). Shiffrin points out that, in contrast, if my acquaintance voiced her reasons for refusal and convinced me to build the shelves without her, then she wouldn't be acting paternalistically (213). What makes the acquaintance's action paternalistic isn't the refusal itself, but the “substitution of judgment and the circumvention of an agent's [my] will” in her refusal (213).

Not only does paternalism not necessarily diminish a subject's freedom, as in the example above, but paternalistic actions may actually enhance a subject's freedom if the subject has explicitly stated that she doesn't want too many choices (Shiffrin 214). That is, it would be paternalistic for my acquaintance to provide me with choices simply because she believes that, despite my desire not to have such choices, I am mistaken in keeping myself in such ignorance or that I would develop a better character by being tested by multiple choices (Shiffrin 214). Similarly, Dworkin points out that such methods as insisting that I hear arguments for my own good may be paternalistic if I want to make my decision impulsively (“Paternalism: Some...” 107). Hearing arguments for my own good may enhance my freedom by expanding my sight beyond my impulsiveness, but, as in Shiffrin's examples, the acquaintance is acting paternalistically because again, she is substituting her judgment for mine regarding how I should develop my agency (Shiffrin 214).

In the above examples, Shiffrin argues that paternalistic action mustn't necessarily diminish a subject's freedom. She then proceeds to argue that neither is paternalism necessarily contrary to a subject's will (Shiffrin 214). Specifically, Shiffrin invites us to suppose that one is

worrying that a subject will, when presented with a certain choice, make the wrong choice. To prevent the subject from making the wrong choice, one preemptively makes the choice for her (Shiffrin 214). For example, a mother might intercept a credit card offer that arrives in the mail for her son and throw it away without telling him in order to prevent her son from becoming the victim of extravagant interest rates (Shiffrin 214). Even if her son doesn't later object to his mother's actions, Shiffrin claims that the mother's actions were nonetheless paternalistic, because once again, the mother substitutes her judgment for her son's regarding what choices are best for him (214). In each of these cases, Shiffrin claims that what makes the action paternalistic are not the effects or expected outcomes of the action (whether they be freedom-diminishing, freedom-enhancing, or even contrary to the subject's will) but rather the motivation behind the action (215). The interference is paternalistic when the interferer substitutes her judgment for the subject's.

Of course, not all cases of substituting one's judgment for another's are cases of paternalism. I am clearly not acting paternalistically, for example, if I ask your advice on which car I should buy but ultimately decide to ignore your advice and buy the car that most pleases me instead. Usually, cases of paternalism involve the interferer's substituting her judgment for the subject's on matters regarding the subject's own welfare because the interferer believes that she knows how to better promote the subject's welfare.

However, Shiffrin argues that this characterization is still overly narrow on two accounts. First, paternalism doesn't necessarily involve a substitution of *judgment*. Rather, one may act paternalistically by thwarting a subject's actions even when the subject's judgment is agreed to be appropriate (Shiffrin 215). For example, it's paternalistic to hide a subject's cigarettes without her consent even if she agrees that she should quit smoking (Shiffrin 215). Thus,

paternalism can be motivated to maneuver around the subject's agency, not just her judgment (Shiffrin 215).

Secondly, Shiffrin doesn't believe that paternalism should be restricted to cases in which the interference is motivated by concern for the subject's welfare (216). Instead, Shiffrin argues that paternalism could involve any cases in which the interferer aims "to take over or control what is properly within the agent's [Shiffrin's 'agent' is who I've been calling the 'subject'] own legitimate domain of judgment or action" (216). The subject's own legitimate domain of judgment or action includes her welfare but is not limited to her welfare. For instance, Shiffrin writes, "Suppose B is the legitimate, duly appointed manager of a group of workers. A substitutes a policy memo that A has written for a memo of B's — not from concern for B's welfare, but from concern for the workers that B manages. A believes that her memo is clearer or establishes a better policy" (216). In this case, even though A is not motivated by concern for B's welfare, Shiffrin claims that she nevertheless acts paternalistically by presuming that her judgment is superior to B's regarding matters that are legitimately within B's domain of judgment and by acting upon this presumption (216).

In summary, Shiffrin's writes that paternalism by A toward B may be characterized as behavior that is "a) Aimed to have (or to avoid) an effect on B or her sphere of legitimate agency; b) That involves the substitution of A's judgment or agency for B's; c) Directed at B's own interests or matters that legitimately lie within B's control; and d) Undertaken on the grounds that compared to B's judgment or agency with respect to those interests or other matters, A regards her judgment or agency to be (or as likely to be), in some respect, superior to B's" (218). As mentioned earlier, Shiffrin arrives at this characterization by evaluating potential cases of paternalism, including those given above, with respect to our "normative reactions" to

paternalism. Thus, rather than characterizing paternalism through specific features — in contrast, Dworkin characterizes paternalistic action by the features of a) being an interference with a subject's liberty or autonomy and b) being motivated or justified by consideration for that subject's own good or welfare — Shiffrin characterizes paternalism based on whether various cases arouse the “normative reactions” we associate with paternalism.

She admits that most accounts of paternalism do not include as broad of a range of behaviors as her own, but questions why we generally draw such a “bright line” between cases in which the interference is motivated by concern for the subject's own welfare as opposed to cases in which the interference is simply motivated by the belief that the interferer's judgment is superior to the subject's regarding some matter within the subject's proper domain of control (Shiffrin 217–218). Both of these cases, Shiffrin claims, “involve the same sort of intrusion into and insult to a person's range of agency,” and this is what we react to as being paternalistic (218). However, Shiffrin doesn't explain what she means by “normative reactions” to paternalism, aside from implying a vague sense of unease or outrage over paternalistic acts. Furthermore, she fails to consider cases in which paternalistic acts might be seen in a positive light (i.e. ones in which the acts might not be considered such an “insult” to a person's range of agency), further obscuring the type of reaction towards paternalism that she expects. As a result, Shiffrin's use of normative reactions as the basis for evaluating whether an act should fall under the term “paternalism” or not is vague and lacks strong argumentative force — after all, if one doesn't share her normative reactions to specific cases, then the points that she draws from these cases are unpersuasive.

Although I do find Shiffrin's characterization compelling in that it appears to capture cases that “seem” paternalistic despite lacking features that are commonly associated with

paternalism, I disagree that cases in which the subject's own welfare is not at stake should be included in paternalism. That is, I agree with Shiffrin to the extent that paternalistic action does not necessarily diminish the subject's freedom, as it may occur through an omission or could even enhance freedom if the subject explicitly asked not to have too many choices. I also agree that one can act paternalistically before the subject is aware of the matter, and so paternalism doesn't necessarily override the subject's will. And finally, I agree that a paternalistic action may be motivated by the belief that the subject's judgment is sound but her ability to *act* is compromised, such as someone addicted to cigarettes who desires to quit but nevertheless continues to smoke.

Shiffrin arrives at these "expansions" to the notion of paternalism by evaluating specific examples, but I don't believe that these features in fact greatly expand the notion of paternalism. All of these paternalistic actions have one feature in common, a feature that was already included in Dworkin's characterization: the paternalistic action interferes with the subject's *autonomy*. Earlier, I adopted Dworkin's characterization of autonomy in which autonomy entails that the subject reflects upon and evaluates her preferences before adopting them as her own — i.e. the subject's preferences are authentic — and is also able to make her preferences effective in her actions. Using this notion of autonomy, actions of omission such as my acquaintance's refusal to help me to build shelves because she believes I would be better off developing those skills on my own interfere with my autonomy; by refusing, without explanation, to assist me in building shelves, my acquaintance prevents me from evaluating my respective preferences (e.g. to be helped versus to learn how to build shelves on my own) and thus I am unable to exercise my autonomy in this case. I am similarly prevented from exercising my autonomy in cases where my acquaintance acts before I am aware of the matter and am able to choose for myself. And if I

explicitly ask not to have too many choices, then I have already made an autonomous decision to limit myself, and so forcing further information on me defies my autonomous decision. Finally, in the case of the cigarette addict, the criteria that autonomy consists not only of the ability to reflect upon and evaluate preferences but also to make preferences effective in action means that interference with the subject's ability to act upon her preference to smoke again thwarts her autonomy. Therefore, Shiffrin's initial expansions to the notion of paternalism may in fact be contained within the feature that paternalistic action interferes with a subject's autonomy, given the notion of autonomy that I have adopted.

On the other hand, I disagree with Shiffrin that interferences motivated by concern for the subject's welfare and those motivated by the belief that the interferer's judgment is superior to the subject's regarding some other matter within the subject's proper domain of control truly involve the *same* sort of intrusion into a person's agency. The subject's own welfare is, to some extent, within her proper domain of control (I won't say that her welfare is *exclusively* within her domain of control, although this may be the case, because I haven't addressed whether others have some legitimate claim on the subject's welfare as well). Even if the subject's welfare does fall within others' proper domains of control as well, I take it as uncontroversial that, at least in cases of autonomous subjects, the subject has a greater claim on controlling her welfare than others. Therefore, to interfere with the subject's judgment or actions on behalf of her welfare is to interfere on a matter that is, to some extent (call it X), within the subject's proper domain of control.

Now consider cases in which the interference isn't motivated by concern for the subject's welfare but by concern for a third party, such as in the example where A substitutes her memo for B's because she believes that her memo is clearer or establishes a better policy. Although

managing the third party may be within B's proper domain of control, I don't believe that the extent to which managing the party is in B's domain of control is the same extent X as before. I'm not sure whether a duly appointed manager has a greater or lesser legitimate say over managing her party than she has over managing her own welfare — that is, I'm not sure whether the extent to which managing the party is within B's domain of control is greater or less than X. Nevertheless, the factors that put her welfare and managing her party within B's domain of control differ significantly. Namely, we tend to suppose that B has a legitimate say over her welfare unless we are given a strong reason not to suppose so (e.g. perhaps B is severely mentally impaired), whereas B is only given legitimate say over her party through others' approval as shown by, e.g., promotion or voting. Because the domain of control is conferred very differently in these cases, it seems reasonable to suppose that the extent to which one's welfare is within one's proper domain of control is a different extent than that to which managing a party is within one's domain of control. Therefore, interferences with judgments and actions regarding these domains of control may constitute different sorts of intrusions into a person's agency.

Interferences with a subject's judgment or actions that are motivated by a concern for her welfare are those typically classified as paternalistic, and indeed what Dworkin would classify as paternalistic as well. In contrast, interfering with a subject's management of a party may be an intrusion into her proper domain of control, but I would not characterize this intrusion as being paternalistic. Moreover, here I am concerned with the conflicting values of autonomy and welfare, and so other potentially "paternalistic" interferences that aren't motivated by concern for the subject's welfare aren't particularly relevant in this context. Therefore, I ultimately adopt the characterization of paternalism given at the beginning of this section. Namely, I will treat

paternalism as an interference with a subject's liberty or autonomy that is motivated exclusively by consideration for that subject's own good or welfare.

Arguments for (at least some) paternalism

Before delving deeper into arguments for paternalism, I wish to make clear that I will not consider arguments from a purely consequentialist account. If one adopts a general consequentialist account, then although autonomy may be included in one's conception of happiness or welfare, it is ultimately one's overall welfare that counts. Thus, on consequentialist views, an act of paternalism is justified if the resulting benefits to the subject's welfare outweigh the harms of overriding her autonomy. As Onora O'Neill writes, "In utilitarian ethical thinking autonomy is of marginal ethical importance, and paternalism only misplaced when it reflects miscalculation of benefits" (173). Because according to most consequentialist accounts, there is only one ultimately value — happiness (or welfare, or some variant thereof) — concern for autonomy cannot actually *conflict* with welfare since autonomy is not valued separately from welfare. In contrast, I assume that autonomy does have value independent of its relation to welfare and thus the two values may conflict. It is within such a view that I wish to examine the possibility of overriding autonomy in favor of welfare through paternalistic interferences.

Hypothetical consent

One way of justifying paternalistic interference is through hypothetical consent. Thus, for example, one may respect autonomy and yet argue that interference is permissible if and only if a choice is irrational and the subject would consent to the paternalist's action if she were fully rational and well informed (Scoccia 318). In taking such a position, Scoccia effectively claims

that a subject's autonomy can be overridden if her actual choice does not conform to her ideal, fully rational choice, where "a choice is rational if it is likely (in light of what the chooser does or can know at the time he makes it) to maximize (within the bounds permitted by his moral principles) the satisfaction of his presently held desires" (Scoccia 320). One can interpret this as Scoccia's prioritizing the subject's welfare over her autonomy. In his case, Scoccia appears to adopt a Desire-Fulfillment theory of welfare, considering his claim that a choice is rational if it is likely to maximize the satisfaction of the subject's presently held desires, which may include having her autonomy respected.⁶ Scoccia claims that it is justifiable to interfere with impetuous choices when the subject holds other, stronger desires that conflict with the impetuous choice. On the other hand, Scoccia argues that this does not necessarily mean that interference with any impetuous choice is justified, for impetuous choices need not be irrational (321), for example when the impetuous choice is likely to maximize the satisfaction of the subject's desires.

Scoccia expands these "desires" to include the subject's values. Presumably he does so because although desires can be fleeting and thus more difficult for others to know, values tend to be more constant and thus it is more likely that others accurately know the subject's values. Indeed, Scoccia implies that not only can others know the subject's values, but they can, to some extent, know the relative weight of the values based on their own value judgments, allowing them to make value judgments on behalf of the subject. Accordingly, e.g., Scoccia believes that one may justifiably interfere with choices based on faulty reasoning such as superstition because the subject may still have other values that align with those of others who are "more scientifically enlightened" (322). Even if the subject's superstitiousness is a stable character trait, the subject may nevertheless endorse other values that conflict with and potentially override

⁶ Although Scoccia writes about "presently held" desires, his argument could be expanded to include past, future, or hypothetical desires.

her superstitiousness. “The mere fact that superstitiousness is a stable character trait in [the subject],” Scoccia writes, “gives us no reason at all to think that his values differ from ours. The mere fact that he continues to prefer laetrile, even after we trot out the proof, also does not show that he has different values. It may only show that he is impossible to reason with” (Scoccia 322–323). Assuming that the subject’s other values *are* similar to ours and that those other values outweigh her superstitiousness, we can presume that she would consent to the paternalistic interference if she were more rational and well informed.

This may appear to assume a lot, but in certain cases, it does not seem so far-fetched to assume that the subject’s values are similar to ours, and that, upon rational reflection, the subject would consent to interference. For example, in the panic of the moment (this will decrease her voluntariness, as described in more detail later; for now, I set aside the issue of voluntariness), the subject may resort to superstition without realizing that her actions put her very survival at risk — yet she probably would, in calmer moments allowing for rational reflection, value survival more than superstitiousness. That said, Scoccia recognizes that the subject may explicitly reject interference, in which case paternalistic interferers would need to recognize her additional value of wanting to act without interference (323) — one might say, her value of autonomy. Most of us, Scoccia admits, have such a value, and so paternalistic interferences are not so commonly justifiable as they may at first appear (323). According to Scoccia, “Our strong desire to run our own lives is quite important for the second view about respect for autonomy because it is the reason why paternalistic intrusion upon irrational acts — acts which do not accurately represent the [subject’s] own values — might still involve an imposition of the intruder’s values on the [subject]” (323).

I find myself wary of Scoccia's claims here. First of all, must we *explicitly* make clear that we value the ability to run our own lives (i.e. our autonomy), or, perhaps, can we assume that people have such a value unless they make the opposite explicit? And in any case, even if people value their autonomy, how do we know how strongly they value it? If they agree with Kant, then presumably they value their autonomy above all else, and so we can know not to interfere. In contrast, Mill, although he was a great champion of autonomy, did allow one exception to his "no paternalistic interference" rule in the case of a subject's attempting to sell herself into slavery. Mill argued that interference is permissible in this case in order to "preserve the liberty of the person to make future choices" (Dworkin, "Paternalism" 118). That is, Mill felt that it was justifiable to interfere with a subject's present autonomy in order to protect her future autonomy. But it is not actually inconsistent for a subject to make an autonomous decision to forfeit her future autonomy, and perhaps she disagrees with Mill's value judgment and instead values her present autonomy more than her future autonomy.

In other words, my biggest worry with Scoccia is that he doesn't explain how or when we can know that a subject's values align with our own and when they don't, and so the boundaries of justified paternalism are unclear. However, it may be possible to use Nussbaum's claim that empirically, people do have some values, needs, and goods in common to determine certain cases in which we can expect a subject's values to align with our own. If we use Nussbaum's capabilities theory of welfare to guide us in determining the subject's values and what would make her better off, then Scoccia's argument can be narrowed and reformulated as follows: it is justifiable to interfere with a subject's choice when that choice does not maximize her welfare, which we can at least partially know based on our theory of welfare. The reason why it is justifiable to interfere with a subject's choice in such a case is because if the subject were fully

rational, she would choose the action that *would* maximize her welfare (namely, the very action that the paternalist wishes to take), which Scoccia takes to imply hypothetical consent.

However, one must be careful about how one uses Nussbaum's theory of welfare in this case. Namely, Nussbaum only defines a limited set of capabilities necessary for welfare.

According to the reformulation of Scoccia's argument, these are the capabilities that a paternalist may be justified in securing for a subject even when doing so is contrary to her choice, because these capabilities represent the values that one can expect to share with the subject. However, if the subject already has the necessary capabilities for a good life, then the further promotion of her welfare is based on her own choices, as Nussbaum does recognize that many values are subjective. Therefore, one could not necessarily refer to one's own value judgments in order to determine a subject's values in areas not regarding her basic capabilities. In such cases, unless the subject has indicated what her values are, one should respect her autonomous choices regarding her life. Indeed, respecting her autonomous choices should be in accordance with the promotion of her welfare.

To give an example in which paternalistic action is justified through hypothetical consent, consider, for example, the case of a badly injured Jehovah's Witness who needs a blood transfusion to survive but refuses treatment on religious grounds. Assume in this case that, in reaching the decision to refuse a blood transfusion, the Jehovah's Witness reflects upon and evaluates the refusal of blood transfusions with regard to her character and concludes that refusing the blood transfusion is indeed in line with her character. Thus, she is authentic. In addition, she is self-determining if she then sets out to determine her life in accordance with her character and judgments — in this case, by refusing the blood transfusion. Being both authentic

and self-determining, then, the Jehovah's Witness chooses to refuse the blood transfusion autonomously.

Without a transfusion, the Jehovah's Witness will die — thus the question is, can paternalistic action of giving the subject a blood transfusion, contrary to her autonomous decision, be justified? Given that one of Nussbaum's basic capabilities necessary for a good life is being able to live the length of a normal human lifespan, the subject's decision to forgo a blood transfusion is contrary to her welfare, despite her decision. The subject cannot have a good life without the capability of being able to live the length of a normal human lifespan (indeed, if she dies she won't have any of the capabilities necessary to have a good life; she won't have a life at all), and by refusing treatment, she gives up the capability to live the length of a normal human lifespan. Therefore, we know that refusing a blood transfusion will not improve the subject's welfare. Moreover, according to Scoccia's claim, if the subject were fully rational, she would choose the action that *would* maximize her welfare, in this case by accepting the blood transfusion. Because the subject would choose the action that would maximize her welfare and we know that the ability to live the length of a normal human life is necessary for a good life, the hypothetical consent of the subject is implied and so paternalistic action to give her a blood transfusion is justified.

Giving a Jehovah's Witness a blood transfusion by appealing to her hypothetical consent despite her actual refusal is problematic, but I argue that the paternalistic action is nevertheless justified. The action is problematic because, by giving the subject a blood transfusion, we override her expressed autonomous decision to refuse treatment. However, I agree with Scoccia that if rational, one would choose the action that would maximize one's welfare. In general, one is the best appraiser of one's welfare — thus, once one's basic capabilities for a good life are

fulfilled, one's welfare is determined by one's own subjective desires and preferences — but there are, as Nussbaum argues, some needs and capabilities necessary for a good life that are shared among us. Where the possession of these capabilities is concerned, we can know a subject's welfare. Therefore, because refusing a transfusion denies the Jehovah's Witness the capability of living the length of a normal human life⁷, we know that she cannot have a good life without a transfusion, and so paternalistic action is justified.

On the other hand, it may be the case that even if the Jehovah's Witness receives the blood transfusion, she still won't have a good life. The possession of the basic capabilities for a good life is necessary, but not sufficient, to guarantee having a good life. Therefore, it's possible that the Jehovah's Witness will end up miserable on account of having received a blood transfusion and so her life doesn't go well even though she survives. The two alternatives in this case are either a) not giving the subject a blood transfusion, in which case she forgoes the capability of living the length of a normal human life and thus cannot have a good life or b) giving the subject a blood transfusion, in which case her life doesn't go well because she is miserable and resents the transfusion. She can't have a good life either way, and so although interfering paternalistically with her choice to refuse a transfusion may make her miserable, such interference doesn't prevent her from having a good life otherwise. However, with a transfusion perhaps she *will* have a good life after all, while without one, we know that she won't. Therefore, overall she will be better off with the blood transfusion, and so the implication of hypothetical consent remains and we are justified in acting paternalistically.

⁷ The exact age that corresponds to “the length of a normal human life” is unclear, although based on actual averages of human lives Nussbaum probably intends it to be around 70 years old. In any case, we can specify that the Jehovah's Witness here is far from that age.

Importantly, the paternalistic interference is only justified in the case of the Jehovah's Witness because we know that refusing the blood transfusion deprives the subject of a capability necessary for a good life and so hypothetical consent is implied. However, paternalistic interference isn't justified in all cases where subjects refuse life-saving medical treatment. In the case of the Jehovah's Witness, by refusing medical treatment the subject deprives herself of the capability of living the length of a normal human life. In contrast, we might consider cases of elderly people — who have already lived “the length of a normal human life” — who refuse life-saving treatment. The elderly subject cannot be deprived of the capability of living the length of a normal human life. When a subject's capabilities necessary for a good life do not come into question, her welfare is instead determined by her own desires and preferences. Therefore, if an elderly subject autonomously refuses life-saving treatment, her decision should — if the decision is indeed autonomous and has thus been informed by her desires and preferences — reflect her welfare. Consequently, hypothetical consent is not implied in this case and so paternalistic interference is unjustified.

Similarly, if the subject has a terminal disease or is otherwise incapable of having good health (another basic capability for Nussbaum) and chooses to forgo life-saving medical treatment, then again paternalistic interference seems unjustified. In this case, the subject need not have lived the length of a normal human life; if her choices are a) to receive treatment and continue living while incapable of having good health or b) to refuse treatment, then under either choice, the subject is deprived of a capability necessary for a good life. Either she loses the capability of living the length of a normal human life if she refuses treatment or she lacks the capability of having good health due to her terminal disease. Nussbaum does not address welfare in cases where the subject cannot have all of the capabilities necessary for a good life, but where

the satisfaction of one basic capability conflicts that of another, it seems that the subject's welfare is generally best served by deferring to the subject's own evaluation of her desires and preferences. Therefore, in cases where a subject's basic capabilities cannot all be satisfied, one should again respect the subject's autonomous choices rather than interfere paternalistically.

Actual consent

Scoccia bases the justifiability of paternalistic interference on the subject's hypothetical consent *if* she were fully rational and well informed. Yet full rationality and information may describe individuals only rarely, and therefore such consent may never, in fact, be given. Instead, Rosemary Carter argues that paternalistic acts are only justified when the subject has alienated her right to autonomy through consent — and to Carter, this does not necessarily mean her hypothetical, fully rational consent. However, Carter does admit that in addition to alienating or forfeiting her right to autonomy through actual prior consent, the subject may forfeit her right through subsequent consent if, subsequent to the paternalistic interference she either explicitly does retrospectively consent or she is “disposed to consent either upon request, or upon the receipt of a relevant piece of information” (136). This subsequent consent must not be caused by the interference itself. That is, the interferer may not brainwash the subject, distort her desires, beliefs, or preferences, or withhold relevant information in order to secure the subject's subsequent consent (Carter 139).

If the subject gives prior consent, then the act may no longer seem paternalistic, for it is not in conflict with the subject's autonomy. However, this is only valid if the consent was given *just prior* to the act. If not, the act, while in accordance with the subject's past autonomy, may in fact conflict with her present autonomy. Presumably in accepting the subject's prior consent,

Carter wishes to protect the subject's ability to ensure that her future decisions reflect her present values, but Carter provides no reason why the subject should have this ability.

To provide a more concrete example: perhaps the subject is currently dieting and, worried that she will break down and gorge herself at a feast later, she decides to give consent to her friend to prevent her from gorging herself. And yet, when the feast comes along, the subject authentically reevaluates her desires and preferences with regard to her character and realizes that no, her earlier desire to diet was not consistent with her character after all and, in fact, she now desires to enjoy her food more than she desires to diet. She then proceeds to act based upon this realization, i.e. she is also self-determining. In short, the subject autonomously changes her mind.

Carter does not seem to consider the dilemma posed here — which should the friend respect, the subject's past or present autonomy? Should a subject even be allowed to forfeit her future autonomy through consent? In a way, forfeiting one's autonomy through consent is a restricted case of Mill's voluntary self-enslavement (where, rather than giving up *all* autonomy, the subject merely relinquishes autonomy over a certain domain of her decisions): it is not inconsistent, perhaps, yet neither should it be respected. If the subject is indeed autonomous at the feast, then her prior consent should be dismissed and her current autonomy respected instead. Perhaps in the past, she did indeed forfeit her consent at the time, but part of being autonomous is the ability to evaluate and change one's judgments accordingly. Thus, if the subject has since changed her mind in a way that preserved her autonomy, then her present choices should be respected over her past ones which have since been rejected⁸.

⁸ If, on the other hand, the subject does not authentically reflect upon and evaluate her desire to eat versus to fast or she doesn't act according to her authentic judgments, then she does not truly

As for subsequent consent, this hasn't actually been given yet, and so it is less clear how the subject's autonomy can truly be forfeited at the time of the interference. Carter offers factors to consider when judging the probability of receiving subsequent consent, such as "whether the paternalistic action is in accordance with the permanent aims and preferences of the subject" (139), but again, the subject may change. Of course, if her aims and preferences are actually *permanent*, then those would not change, but permanence, although possible, cannot be guaranteed ahead of time. As Husak points out, "even a reasonable belief that consent will be given by an agent does not guarantee that his autonomy will not be violated, for autonomy may allow an agent to choose to do what others could not have reasonably expected" (32). Even in cases where it seems extremely likely that the subject's aims and preferences are permanent, the subject's autonomous choices *now* are even more certain because they can be explicitly communicated. Therefore, it is still preferable to respect the subject's current autonomous choices.

Thus, Carter's conception of consent fails to provide justifiable grounds for paternalistic interference. However, while it does not seem so justifiable to refuse to respect a subject's present decisions on the basis that her autonomy is forfeit through prior or subsequent consent, one may perhaps interfere on the basis that her present autonomy is compromised. If this is the case, then the subject is unable to autonomously decide for herself anyway, and so we may be justified in stepping in to help. This autonomy may be compromised if the subject's decision is involuntary or if she is incompetent.

act autonomously when gorging herself. In this case, she does not currently have autonomy to respect and so her past, autonomous choice should be respected instead.

Involuntariness

According to Feinberg, “one assumes a risk in a fully voluntary way when one shoulders it while fully informed of all relevant facts and contingencies, with one’s eyes wide open, so to speak, in the absence of all coercive pressure. There must be calmness and deliberateness, no distracting or unsettling emotions, no neurotic compulsion, no misunderstanding” (“Legal...” 7). Thus, compulsion, misinformation, excitement or impetuosity, clouded judgment, or immature or defective faculties of reasoning may all diminish the voluntariness of a decision (Feinberg, “Legal...” 7). Consequently, voluntariness admits of degrees. Feinberg argues that although Mill wanted to prohibit paternalistic interferences with fully voluntary choices and probably also with “almost-but-not-quite fully voluntary choices” and even “some substantially nonvoluntary ones,” he certainly did not hold such qualms about interference with completely involuntary choices, for “insofar as the choices are not voluntary they are just as alien to the individual as the choices of someone else” (Feinberg, “Legal...” 8). Respecting a subject’s involuntary choice does not actually respect her autonomy, for an involuntary choice is not an autonomous one.

Furthermore, a paternalistic act is justified if it is necessary to determine whether the subject is acting voluntarily. For example, one would be justified in seizing the subject to prevent her from crossing an unsafe bridge because, given that most people do not actually want to cross unsafe bridges, it is safe to assume, at least on a temporary basis, that the subject is currently acting involuntarily due to misinformation or ignorance (Feinberg, “Legal...” 8). However, if the subject then verifies that she is, as a matter of fact, aware of the bridge’s unsafeness, then her action is no longer involuntary and, since her autonomy isn’t compromised after all — assuming that it isn’t compromised through other factors such as clouded judgment

— one is no longer justified in interfering (Feinberg, “Legal...” 8). Of course, in this case it seems possible that the subject’s autonomy *is* still compromised by, e.g., clouded judgment or compulsion, and so Feinberg suggests that in such cases, one is justified in gathering proof or evidence that the subject is truly acting voluntarily before allowing her to cross the unsafe bridge (“Legal...” 9). This evidence of voluntariness may include “elaborate tests, swearings, psychiatric testifying, waiting periods, public witnessing, and the like” (Feinberg, “Legal...” 12).

Arneson objects that Feinberg’s definition of “fully voluntary” is too narrow and almost equivalent to “fully rational,” which, he argues, doesn’t accurately reflect Mill’s thinking after all (Arneson 484). Instead, Arneson claims that Feinberg seems to commit himself to “the distinctly un-Millian position that all acts involving mistakes are nonvoluntary and as such fall beyond the protected scope of the antipaternalism principle” (485). Yet, as stated above, Feinberg explicitly writes that although Mill probably wanted to prohibit paternalistic interferences with fully voluntary choices and even some nonvoluntary ones as well, he did not want to prohibit interferences with *completely involuntary* choices. Although Feinberg’s suggestion that the subject provide evidence to “prove” her voluntariness seems to imply that he would like to ensure that the subject’s action is fairly voluntary, he could lower his standards of proof and evidence to the point of ensuring that, at the very least, the subject isn’t acting completely involuntarily. This, Arneson may be willing to agree, is more “Millian.” So, it seems at least justifiable to interfere in what appear to be completely involuntary decisions until one discovers otherwise.

However, in addition to paternalistically intervening with completely involuntary decisions, it seems that paternalistic intervention in decisions that are mostly involuntary should also be permitted. In order to determine what this means, I propose a somewhat less stringent

view of voluntariness. Rather than requiring the subject to be “fully informed of all relevant facts and contingencies,” as Feinberg does, I would like to first consider the relationship between autonomy and voluntariness. Earlier, I adopted Dworkin’s view of autonomy, according to which autonomy consists of both authenticity and self-determination, where authenticity refers to the reflection upon and evaluation of one’s preferences and judgments as well as the ability to alter these preferences based on one’s evaluations, while self-determination refers to being able to determine one’s life in accordance with one’s character and judgments. Voluntariness implies that one is choosing one’s course of action based on one’s considered judgments; thus, it seems that voluntariness includes self-determination. If one’s actions do not reflect one’s judgments, then the action is not voluntary and so paternalism is justified. However, voluntariness does not seem to necessarily imply a strong sense of authenticity. For example, I may form judgments without a significant amount of reflection, but while acting upon these judgments may be rash and inadvisable, it’s not necessarily involuntary. Thus, despite the fact that I am generally a cautious person, I choose on a whim to go bungee-jumping. Although I may not be perfectly autonomous at this point because I did not sufficiently evaluate my desire to go bungee-jumping and indeed it is contrary to my character, I may nevertheless be perfectly aware of what I am choosing to do and am acting voluntarily.

Secondly, I believe that voluntariness implies a sense of what someone “wants,” meaning that the subject’s desires also come into play when determining the extent of her voluntariness. Perhaps, then, we should also appeal to the subject’s strongest desires, as we did in Preference-Hedonism. Consider again the case of the subject crossing the bridge. Perhaps she does presently desire to cross the unsafe bridge, but another desire, e.g. a hypothetical desire to *safely* reach the other side or a future desire to do something which would be impossible were she to

severely injure herself or die in the present, might be stronger than her present desire. If the subject has other, stronger desires than the one presently compelling her to cross the bridge, it doesn't seem that she really "wants" to cross this bridge, all things considered. Therefore, even if she is presently self-determining, this is a case in which I would deem the subject mostly involuntary, and consequently paternalism is appropriate here. On the other hand, if the subject has other desires that conflict with her current one but which are weaker, then the subject is, despite these reservations, mostly voluntary, and so paternalism is not necessarily justified.

Incompetence

A second way in which autonomy may be compromised is through incompetence. Competency, according to Christman, "includes various capacities for rational thought, self-control, and freedom from debilitating pathologies, systematic self-deception, and so on." Carter argues that a subject's action rights (meaning the rights to do or refrain from doing something, such as walking down the street at night) are compromised if she lacks the relevant capacities (such as the capacity to appreciate the consequences of the proposed actions) (143). Therefore, in cases of incompetence, there is no need to respect the subject's right to act upon her decisions because her rights are compromised anyway.

However, defining cases of incompetence is no easy task. Typically, one is considered "competent" based on a threshold standard. That is, one is "competent_t," as Brock puts it, if one meets a certain threshold for competence, and one is incompetent if one fails to meet that threshold (Brock, "Paternalism and Promoting..." 241). Using this conception of competence, we can justify the view that it is acceptable for normal persons to act paternalistically toward the mildly retarded, who are considered incompetent_t, while it is unacceptable for the gifted or very

intelligent to act paternalistically toward normal persons, who are considered competent_t (Brock, “Paternalism and Promoting...” 241 – 242).

But how do we set the threshold for competence? Carter considers the class of incompetents to be “those who are unable to understand or practice satisfactorily the basic requirements of survival, and so whose lives would be at worst in constant peril, and at best grossly unhappy, if it were not for intervention on the part of others” (143). Based on this definition, it seems that Carter’s threshold for competence is the ability to survive. And indeed, we probably do want to interfere with the decisions of subjects that are unable to even survive without help. Maybe this is all the interference we wish to justify. But maybe we would rather not set the threshold for competence quite so low, in which case we returned to the dilemma of setting an appropriate and non-arbitrary threshold for competence.

Conclusion

Of the justifications for paternalism considered here, only Scoccia’s hypothetical consent seems to justify paternalism in cases where the subject’s autonomy is not currently compromised. By appealing to the subject’s hypothetical consent, we may override the subject’s autonomous decision if we know what will make the subject better off. Yet because the subject’s welfare is in large part dependent on her own desires, our knowledge of the subject’s welfare is limited. Accordingly, I refer to Nussbaum’s capabilities theory of welfare to argue that at least in cases where the subject’s autonomous decision will deprive her of basic capabilities necessary for a good life or will keep her in a position lacking such capabilities, interference with a subject’s action that gives her these capabilities will promote her welfare. Thus, for example, we are justified in overriding a Jehovah’s Witness’s autonomous refusal to

receive a life-saving blood transfusion because such a refusal would deprive the Jehovah's Witness of the capability of living the length of a normal human life. On the other hand, if the subject already has the basic capabilities necessary for a good life, or if it is impossible for her to have all of these capabilities (e.g. if, regardless of how she chooses, she will either lack the capability of living the length of a normal human life or of having good health), then we no longer have clear knowledge of how to promote the subject's welfare. Without such knowledge, we cannot rely on the subject's hypothetical consent, and so paternalistic interference is not justified.

Where appealing to hypothetical consent is not possible, other arguments for paternalism fail to provide justification for paternalism when the subject's autonomy is intact. If the subject's autonomy is currently intact, Carter's arguments for justifying paternalistic interference through prior or subsequent consent are unconvincing because it is unclear why we should respect a subject's prior or subsequent consent if she is *now* autonomous as well. On the other hand, if a subject's prior or subsequent consent is given and her autonomy is currently compromised, then the issue of respecting her current autonomy doesn't come into play because there is no current autonomy to respect. However, in cases where the subject's autonomy is compromised, the paternalistic interference is justified more on account of the subject's autonomy being compromised, such as through involuntariness or incompetence, than on account of the subject's prior or subsequent consent.

So ultimately, there are only a very limited number of cases in which paternalistic interference is justified in overriding an autonomous subject's decision. This may be frustrating for those who wish to aid an autonomous subject that seems to be making a terrible decision, but this is the price we pay for respecting autonomy. And indeed, we ourselves would generally

wish our own autonomy to be respected as well. To determine, intuitively, just how important one's autonomy is to oneself, Feinberg suggests putting oneself in the shoes of the subject (*Harm...* 62). The subject's autonomous decisions, presumably, are based on the belief that they reflect her interests or, at the very least, aren't bad for her (if she did genuinely believe she was making a terrible decision, Feinberg claims that "that would be so irrational that it would put the voluntariness of [her] choice in doubt"). "If [s]he chose that alternative because [s]he believed it good (or at least not bad) for [her]self," Feinberg continues, "then either the difference between [her] and [her] would-be constrainters is over some matter of fact about which [s]he is simply mistaken, in which case [s]he would welcome being set right, or it is about the nature of [her] self-interest, or the reasonableness, given [her] values, of the risks [s]he wishes to assume" (*Harm...* 62). If the difference concerns the nature of her self-interests or the reasonableness of the risks she wishes to assume, the "disagreement would be more intractable," and the subject would resent having her autonomous decision overruled in favor of others values (Feinberg, *Harm...* 62). After all, it is her interests, her risks — and thus, more than anyone else, she should have ultimate say in her decision.

Works Cited

- Arneson, Richard J. "Mill versus Paternalism." *Ethics*. 90.4 (1980): 470-489. Print.
- Brock, Dan. "Paternalism and Promoting the Good." *Paternalism*. Ed. Rolf Sartorius. 1st ed. Minneapolis: University of Minnesota Press, 1983. 237-260. Print.
- Carter, Rosemary. "Justifying Paternalism." *Canadian Journal of Philosophy*. VII.1 (1977): 133-145. Print.
- Christman, John. "Autonomy in Moral and Political Philosophy." *Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, 2010. Web. 19 Dec 2011.
<<http://plato.stanford.edu/entries/autonomy-moral/>>.
- Dworkin, Gerald. "Paternalism." *Morality and the Law*. Ed. Richard A. Wasserstrom. 1st ed. Belmont: Wadsworth Publishing Co., 1971. 107-26. Print.
- Dworkin, Gerald. "Paternalism: Some Second Thoughts." *Paternalism*. Ed. Rolf Sartorius. 1st ed. Minneapolis: University of Minnesota Press, 1983. 105-112. Print.
- Dworkin, Gerald. *The Theory and Practice of Autonomy*. 1st ed. New York: Cambridge University Press, 1988. Print.
- Feinberg, Joel. "Legal Paternalism." *Paternalism*. Ed. Rolf Sartorius. 1st ed. Minneapolis: University of Minnesota Press, 1983. 3-18. Print.
- Feinberg, Joel. *Harm to Self*. 1st ed. New York: Oxford University Press, 1986. Print.
- Herzog, Don. *Happy Slaves: A Critique of Consent Theory*. 1st ed. Chicago: The University of Chicago Press, 1989. Print.
- Husak, Douglas N. "Paternalism and Autonomy." *Philosophy & Public Affairs*. 10.1 (1981): 27-46. Print.

Nussbaum, Martha C., and Jonathan Glover. "Women, Culture, and Development." *Oxford Scholarship Online*. (2003): 61-104. Print.

O'Neill, Onora. "Paternalism and Partial Autonomy." *Journal of Medical Ethics*. 10.4 (1984): 173-178. Print.

Parfit, Derek. *Reasons and Persons*. 1st ed. Oxford: Clarendon Press, 1984. Print.

Scoccia, Danny. "Paternalism and Respect for Autonomy." *Ethics*. 100.2 (1990): 318-334. Print.

Shiffrin, Seana Valentine. "Paternalism, Unconscionability Doctrine, and Accommodation." *Philosophy & Public Affairs*. 29.3 (2000): 205-250. Print.

Sunstein, Cass R. "Preferences and Politics." *Philosophy & Public Affairs*. 20.1 (1991): 3-34. Print.