

SELF-DETERMINATION AND MORAL RESPONSIBILITY

by

EZRA FRIED

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2010

© 2010

EZRA FRIED

All Rights Reserved

This manuscript has been read and accepted for the Graduate
Faculty in Philosophy in satisfaction of the dissertation
requirement for the degree of Doctor of Philosophy

Stefan B. H. Baumrin

Date

Chair of Examining Committee

Iakovos Vasiliou

Date

Executive Officer

Peter Simpson

Michael Levin

Stephen Grover

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SELF-DETERMINATION AND MORAL RESPONSIBILITY

by

Ezra Fried

Advisor: Professor Peter Simpson

“Self-Determinism” is the claim that we can originate acts – that someone can be the first source of his act. It is often thought that freedom and moral responsibility require the ability to originate acts. I argue that this is not so. However, there is a special kind of moral responsibility that we can have for an act only if we have originated it or might have originated another. Someone has this “pure” moral responsibility for his act just in case he deserves that we take the purely moral attitude toward him of *liking or disliking him as a person* in considering the act, as opposed to our also considering and explaining his performance of it in an objective, scientific way. I argue that the origination of an act can be understood in a way that preserves the dominant view of causation according to which events (including acts) are always caused by other events. A beginning-less series of originative acts terminates in the final originated act. This is an origination because the agent is the subject of every act in the series. I advocate a moderate, moral self-determinism according to which an agent’s personality restricts the range of acts that he can originate, and according to which he originates them only for the sake of their rightness.

Table of Contents

Introduction – 1

1. Action, Intention, and Desire – 8

2. Freedom and moral responsibility – 36

3. Moral goodness and badness – 68

4. An event-causal account of origination – 87

5. Moral self-determinism – 106

6. Pure moral responsibility – 124

7. Moderate moral self-determinism and behavioral probability – 140

8. Previous self-determinist discussions of behavioral probability – 169

9. Summary of the proposal, and discussion of relevant issues – 191

Bibliography – 220

Introduction

People often think that we cannot be morally responsible for an act unless we have originated it or might have originated another. To “originate” an act is to be its first source. Someone’s origination of an act is usually understood as his causing it, where nothing causes him to cause it. This understanding of origination relies upon the notion that a substance can cause an event: an agent is a kind of substance, and his act is the event that he causes.

I will argue we *can* be morally responsible for an act that we have not originated, even if we could not have originated another. However, I will also argue that there is a special kind of moral responsibility, which I call “pure” moral responsibility, which we can have for an act only if we have originated it or might have originated another. Someone has pure moral responsibility for his act just in case he deserves that we take the purely moral attitude toward him of *liking or disliking him as a person* in considering the act, as opposed to our also considering and explaining his performance of it in an objective, scientific way.

I will propose a model of origination that does not rely upon the notion that a substance can cause an event. I will propose that origination can be understood in a way that preserves the presently dominant view of causation according to which events (including acts) are always caused by other events.

I call the proposal that we can originate acts “self-determinism”. According to an *extreme* self-determinism, someone’s personality (his tendency to have certain desires of certain strengths in certain circumstances) cannot restrict the range of acts that he can originate. This seems incompatible with the evident rough predictability of our behavior.

Childhood abuse makes someone more likely to behave violently as an adult. This strongly suggests that the childhood abuse has given him a tendency to have violent desires that *does* restrict his ability to behave nonviolently. I will therefore propose a *moderate* self-determinism, according to which someone's personality restricts the range of acts that he can originate. He can originate acts only within this restricted range. Moreover, I will propose a moderate *moral* self-determinism, according to which someone can originate acts only for the sake of what he takes to be their greater *rightness*, as compared to the acts that he is inclined to originate.

According to my proposal, we originate an act by originating a desire of limited strength to do whatever is right, or by originating the strengthening of an existing desire to do whatever is right. This alters the overall balance of our desires, inclining us to do something that we hadn't been inclined to do – something that seems to us to be more right than what we had been inclined to do. Our new inclination then in turn causes the emergence in us of a corresponding intention, which causes the intended behavior.

According to my "event causation" picture of origination (in which events are always caused by events, rather than by a substance such as an agent), an originated act (or intention) is the final term of a series of events, each caused by the one preceding it, that has no beginning. Suppose someone originates the keeping of a promise. This act is caused by his intention to keep the promise, which is caused by the strengthening of his desire to do whatever is right, which is caused by his origination. This origination is caused by his intention to perform it, which is caused by a prior origination, which is caused by a prior intention to perform this prior origination, which is in turn caused by a prior intention, and so on. This beginning-less series of origination acts (and

intentions) constitutes an *origination* of the final, promise-keeping act because the agent is the *subject* of every act in the series. It may seem preposterous that this infinite series of events should occur in the very finite periods of time during which we make morally significant decisions, without our being distinctly aware of each individual act in the series. The origination acts all occur simultaneously, however, and since all of the origination acts in the series are exactly similar (as are the corresponding intentions) they merge with one another in our introspective view of them. We feel the whole series as a single event of deciding.

I will argue that this proposal's preservation of what I call "pure" moral responsibility should satisfy those who insist upon attributing this sort of moral responsibility to us, despite its great restriction of the range of intentions and acts for which we can have this pure moral responsibility.

In the first chapter, I sketch a picture of action as behavior caused by intention. I propose that someone's intention to do something is caused by his inclination to do it, which is a function of the overall balance of his desires, together with his beliefs regarding how these desires may most effectively be satisfied. Someone may be inclined to do something that he doesn't want (have a desire) to do, either as a means to an end that he does desire, or as a compromise between competing desires. I emphasize that desires are events or states of which a person is the subject, rather than entities inhabiting or afflicting him. I defend the view that desires (or their onsets) are forceful, tending to cause acts by generating intentions. I argue for the superiority of my intention-based picture of action to two alternate pictures of action. According to one of these, endorsed by Carl Ginet and Lawrence H. Davis, action is distinguished by the presence of a

“volition.” According to the other, endorsed by Richard Taylor and Roderick Chisholm, action is behavior caused by an agent.

In the second chapter, I offer a definition of moral responsibility and explain its terms. This definition of moral responsibility depends upon the thought that when we reasonably attribute certain properties to someone we should like or dislike him *as a person* in considering his act. If he really has these properties, he is morally responsible for his act. I argue that we can act freely without being morally responsible for our acts. I argue that origination ability is required neither for freedom nor for moral responsibility. I consider and reject non-self-determinist proposals by Harry Frankfurt, Gary Watson, and John Martin Fischer regarding what freedom and moral responsibility require. I endorse Susan Wolf’s proposal that moral responsibility requires an ability to form reasonable beliefs about the rightness and wrongness of possible acts. I propose as an additional requirement that when someone is morally responsible for his act his intention must express his personality. To act freely is to act from a personality-expressing intention; thus freedom is one of the requirements for moral responsibility but is not sufficient for it.

In the third chapter, I develop an account of moral goodness and badness. Some such account is required by the proposed definition of moral responsibility, since it relies upon the notion that we should sometimes like or dislike someone as a person in considering his act. We should like or dislike him as a person only if we reasonably believe him to be morally good or bad in acting as he does. Additional motivation for offering an account of moral goodness and badness is provided by the consideration that it is chiefly because of the importance that many self-determinists give to (moral) *credit*

and *blame* that they persist in their self-determinism; they think that credit and blame are justly given only when the person in question has originated his act or might have originated another. Someone deserves credit or blame only insofar as he is morally good or bad. Loosely following D. H. Prichard and W. D. Ross, I propose that there are three kinds of moral goodness: conscientiousness, idealism and virtue. Someone is conscientious in acting for the sake of the act's rightness, idealistic in acting for the sake of the act's superlative goodness, and virtuous in doing for its own sake what is in fact best. Idealism and virtue require that an act not be performed if it is thought wrong, and so involve an overriding conscientiousness.

In the fourth chapter, I present the "event-causal" account of act-origination that I summarize above, and argue for its superiority to an account picturing an agent as causing his own act. In the fifth chapter, I distinguish moral from amoral self-determinism. According to moral self-determinism, we originate acts only for the sake of what we take to be their greater rightness, as compared to those we are inclined to perform. We are therefore unable to originate acts that we don't regard as more right than those we are inclined to perform. Following C. A. Campbell, I argue for the superiority of moral self-determinism on introspective grounds. We are justified in believing what we can't help believing. Moral deliberation (over whether to follow inclination or do the right thing) requires that we believe that we can originate doing the right thing. So, we are justified in believing that we can originate acts for the sake of their rightness. However, because of the strangeness of the notion of an originitive ability we shouldn't believe that we can originate acts unless we have to believe this. We never find it necessary to believe that we can originate acts in circumstances other than

moral deliberation. So, we are not justified in believing that we can originate acts for the sake of anything other than their rightness.

In the sixth chapter, I distinguish “ordinary” moral responsibility from the “pure” moral responsibility that would require an ability to originate acts. I consider and reject two alternate pictures of the special sort of moral responsibility that the origination of an act or the ability to originate another would incur: a “deep” as opposed to “shallow” moral responsibility, and an “eternal” as opposed to “mundane” moral responsibility. I consider certain problem cases: for example, how can we justify *not* thinking of someone as deserving blame for refraining from doing something that he unreasonably and falsely thinks right (the case of Huckleberry Finn), while at the same time thinking of a perversely conscientious tyrant as deserving pure blame for originated acts that he unreasonably and false thinks right? I speculatively retrieve the notion of “eternal” moral responsibility as available to those who require some way to distinguish between more and less intense grades of special blameworthiness incurred through the perversely conscientious origination of bad acts.

In the seventh chapter, I distinguish moderate from extreme self-determinism; both may be either moral or amoral. I explore the notion of willpower, proposing that willpower be understood as an ability to originate desire. I argue that the rough predictability of our behavior gives us reason to prefer a moderate self-determinism to an extreme one. I argue that moderate moral self-determinism sufficiently meets the same psychological needs that may seem to be met only by extreme moral self-determinism. We want to feel that we possess and exercise maximal power, and that our normal punitive and reward-bestowing practices are just. We want to avoid the apparently

paralyzing implications of the view that when we engage in everyday deliberation we often don't originate our decisions. We want to believe that we are enduring, continuous selves genuinely involved in the world. We satisfy these desires as moderate moral self-determinists. Since a moral self-determinism is to be preferred to an amoral one on the introspective grounds examined in the fifth chapter, a moderate moral self-determinism turns out to be preferable overall.

In the eighth chapter, I examine ways in which extreme moral and amoral self-determinists including Randolphe Clarke, Timothy O'Connor, Robert Kane, Laura Ekstrom, C. A. Campbell, and Thomas Reid have previously addressed the apparent incompatibility of their position with the predictability of our behavior. I find their approaches to be inadequate. I note a moderate amoral self-determinist proposal considered in passing by Timothy O'Connor, and acknowledge C. S. Lewis's casual suggestion of what looks like a genuine moderate moral self-determinist proposal.

In the ninth chapter, I offer a formal outline of my proposal, and discuss relevant issues. These include the problem of how it can allow specific sorts of moral improvement; how it can deal with cases in which someone seems to force himself to do what he thinks less right than what he is inclined to do and cases in which someone seems to aim at a lesser wrongness rather than at a greater rightness; how there can be unsuccessful attempts to originate an intention; how it can deal with the problem of weak will and with moral compromises; how it can deal with cases of natural saintliness, when the ability to originate an act for the sake of its rightness would seem to be precluded; and how a way of life might be originated.

1. Action, Intention, and Desire

If someone is to be morally responsible for what he does, he must *intentionally* do what he does. To say that someone intentionally does what he does is to allow that someone might do something unintentionally, just as rocks, planets, and rain do things unintentionally – rolling, orbiting, and falling, for example. Of course, when we say that someone does something, or behaves in a certain way – I will use the words “do” and “behave”, and their various derivatives interchangeable – we usually have in mind that they do it intentionally. But in saying that someone does something intentionally we abandon this assumption. It will be helpful, in developing an account of moral responsibility, to use the words “do” and “behave” in the latter way, isolating someone’s behavior from the intention that renders it intentional, so that we can describe what someone does without referring to this intention.

But what is it that someone does, which he might or might not do intentionally? Shall we say that what he does includes the effects of the motions of his body and mind, along with these motions themselves, when these effects are events of which he is not the subject? In this case, to use Donald Davidson’s example, when Hamlet thrusts forward his hand, which is closed around the handle of a rapier, with the intention of thereby bringing about the death of “the man behind the arras”, who turns out to be Polonius, the killing of Polonius would be something that Hamlet does.¹ Or, shall we say that what someone does is limited to motions of his own body and mind? (Thinking, imagining, feeling, and deciding would be “motions” of someone’s mind.) In this case, Hamlet’s doing would be restricted to the thrusting of his arm forward in a certain direction with his fingers clenched around the handle of the rapier, together with whatever is going on in

¹ Donald Davidson, “Agency”, in *Essays on Actions & Events* (Clarendon Press: Oxford, 1980), 43-61, 46.

his mind at the time. In ordinary conversation it would be very strange to deny that “he killed Polonius” is a correct answer to the question, “What did Hamlet do?” However, it will again be helpful in developing a clear account of moral responsibility to follow Davidson in distinguishing between someone’s “core” doing or behavior (I will use these terms interchangeably), and his “greater” doing or behavior.² Someone’s core behavior just consists in the motions of his own body and mind. His greater behavior consists in the bringing about of further events by this core behavior. (Davidson sees only bodily events as belonging to the core act; I add mental events because it seems to me that I can think about something intentionally, directing my thoughts toward this subject, and that in this case my thinking is an act.)

When we speak of someone as doing something, or as behaving in a certain way, we ordinarily have his greater behavior in mind. This more ordinary use of the words “doing” and “behavior” is also looser, because the effects of someone’s core behavior ripple outward from it indefinitely. Hamlet’s core behavior not only causes the death of Polonius; it also causes the grieving of Ophelia; Hamlet makes Ophelia grieve. When I have someone’s core behavior in mind I will say so; otherwise I will be using the word “behavior” to designate someone’s greater behavior.

I will use the word “act” to indicate someone’s doing something, or behaving in a certain way, when he does it, or behaves in this way, intentionally. Just as I distinguish between someone’s “core” behavior and his “greater” behavior, so I will distinguish between someone’s core act and his greater act. Someone’s core act is his intentional core behavior. Someone’s greater act is his intentional greater “behavior”. Hamlet performs the greater act of killing a man behind an arras because he intends to bring

² Ibid., 59. Davidson uses the adjective “primitive”, rather than “core”.

about this man's death by means of his core act of thrusting his arm in the direction of the arras with his fingers clenched around the knife-handle. He doesn't perform the act of killing Polonius though, unless he knows that the man behind the arras is Polonius. (I allow that killing Polonius is something that Hamlet *does*. After all, he kills Polonius. Killing Polonius just isn't an act that he performs.)

Ordinary speech permits our using the words "act" and its derivatives without any reference to intentions, just as it permits our using the words "do" and "behave" with an implied reference to intentions. Thus, we might speak of the action of glaciers upon landscapes, just as we speak of someone's immoral behavior. However, in ordinary speech we rarely say that things that are incapable of intending perform acts. When we say this, we usually do so in an implicitly metaphorical, and often humorous, manner. This is enough to justify distinguishing between mere doing or behaving, on the one hand, as not requiring any intention, and acting, on the other hand, as consisting in intentional doing or behaving.

In using the word "act" to designate someone's intentional behavior, I use it to designate his *intended* behavior. So, for example (taking "act" in its loose, ordinary sense rather than its "core" sense) since Hamlet intends to bring about the death of the man behind the arras with his rapier-thrust, but not the death of Polonius, he performs the act of killing a man behind an arras but not the act of killing Polonius. Since this Hamlet-example is taken from Donald Davidson, it should be noted that Donaldson does not agree that someone's act (in the loose, ordinary sense of "act) consists only in the behavior that he intends to perform. He instead proposes that "a man is the agent of an

act if what he does can be described under an aspect that makes it intentional”.³ Since Hamlet’s bringing about of Polonius’s death can be described as his killing of the man behind the arras, which is something that he does (in the loose, ordinary sense of “do”) intentionally, his bringing about of Polonius’s death is an act. As soon as we allow that Hamlet’s bringing about of Polonius’s death is Hamlet’s act, we might as well describe this act as his killing of Polonius. So, Hamlet performs the act of killing Polonius.

Here, Davidson only reports our ordinary way of using the word “act”. It would certainly be extraordinary to claim that Hamlet *doesn’t* perform the act of killing Polonius. After all, if Claudius (ignorant of what had transpired) were to ask an observer how Hamlet had “acted” that day, the observer would likely say that Hamlet had killed Polonius. Still, my restriction of the word “action” to designate only what the agent intends to do is justified by its usefulness. It enables me to develop a neat account of moral responsibility in terms of favorable or unfavorable attitudes that should be taken toward people in response to their acts. It seems to me that when we should respond to people’s acts by assuming favorable or unfavorable attitudes toward them, these responses should be to acts described only in terms of what their agents intended to do. Note that I don’t *diverge* from ordinary usage in defining action as I do; rather, I am taking a part of ordinary usage and treating it as the whole. My stipulation that an act is intentional, meaning *intended*, behavior states what an act is *from the point of view of the agent, before he performs it*.

A core act has nearer and further consequences. Someone’s greater act is his intended bringing about of one or more of these consequences through the performance of his core behavior (which is also his core act, since he intends it). We can say that

³ Ibid. 46.

someone's "ultimate goal" in performing his core act is the event or events that he performs his core act in order to bring about. Or, we can say that this is what he ultimately intends to accomplish by his core act. Suppose that Hamlet in fact knows that the man behind the arras is Polonius. In this case, perhaps Polonius's death is Hamlet's ultimate goal in rapidly extending his arm, etc. Alternately, Hamlet's ultimate goal might be the collapse of Polonius's business. (Hamlet is a rival entrepreneur.) Hamlet doesn't particularly care whether or not Polonius dies; what interests him is the collapse of Polonius's business. Or, Hamlet might have simultaneous ultimate goals in performing his core act, one being Polonius's death and another being the collapse of Polonius's business.

What happens when someone acts intentionally? He has an intention, and he does something, but what is the connection between his intending and his doing, such that his doing is an intentional doing, an act? The prevailing view among theorists of intention is that it is a state, not a continuing event. If we accept this view, and also accept the principle that events, not states, cause things to happen, then we might hold that the onset of an intention to do something at the present time causes someone to begin doing it, and that he must then continue to do it as long as his intention persists. He "must" continue to do it as long as his intention persists for the same reason that water must boil as long as it is in the state of being very hot – a law of nature is in effect. Simultaneous intentions might then arise to restrict the manner in which someone does what he intends to do; otherwise, he will simply do it as efficiently as possible. For example, someone intends to go somewhere now, and the onset of this intention causes him to begin going there, but on his way there, when shoving aside someone who is in his path would get him there

most efficiently, the onset of a simultaneous intention to walk around the person causes him to do so.

Obviously, during most of our activity we are not intending anything, even though this activity is intentional. We are on “autopilot.” But this activity is still caused by the onset of an intention that was present when it began. Or, if an act is linked together with several other acts in a habitually performed sequence, the whole sequence is caused by the onset of an intention that ceases to be felt, and therefore (I would assume) ceases to exist once the activity is underway. The intention revives only when some obstacle arises, as when someone gets in one’s way when one is hurrying somewhere.

Although I defer to the prevailing view that the *onset*, rather than the continuation, of a pro-attitude would have to be what causes behavior, I can’t help thinking, though, that someone’s intention to do something now really does continuously cause his doing of it. Since I accept the principle that events, not states, cause events, this leads me to think that intentions are continuing events. This way of thinking is supported by our ability to say that someone *intends* to do something. Similarly, what might be called the “on”-state of a car’s engine is really just a matter of the engine’s *running* – a continuing event that causes the car to move. Of course, the running of an engine consists in thousands of successive explosions, each one an event in its own right. But something analogous might be happening where persistent intentions are concerned. Even if this isn’t so, it can be said that someone’s intention to do something now causes his doing of it, rather than that the onset of the intention causes the doing. The shorter formula can be regarded as a convenient abbreviation whenever it appears. The reader

who is more comfortable than I am with this picture can just translate the shorter formula into the longer one.

I will ignore intentions to do things in the future, so as not to have to deal with the question of how these are causally related to the doing of those things. In speaking of intentions, I will be referring to intentions to do things now – “present intentions.” It seems to me that intentions with respect to future doings – “future intentions” – would in any case have to undergo some sort of transformation in order to become present intentions, when the intended time of doing arrives, if they are to play a role in the causing of behaving. This transformation would be the onset of a present intention.

It might be urged that we only intend to *try* to do things. But it seems to me that to intend to try to do A is really only to intend to do A *and* to intend to do B as a means to doing A. One’s intention to do B is then one’s intention to try to do A. There are hierarchies of intentions, ordered in accordance with the means-to-ends ordering of the intended deeds. If what someone intends to do is his ultimate end, then he wants to do what he intends to do. Someone can also have what he intends to do as an ultimate end, thus wanting it, while at the same time intending to do it as a means to something else that he also wants.

Shall we call someone’s act the combination of his intention to behave in a certain way and the behavior that his intention to behave in this way causes, or apply the word “act” to the behavior alone, given that it is caused by his intention to behave in that way? This is not a substantive issue; we only need a verbal stipulation. It is more convenient to call the behavior itself an “act”. Moreover, this seems to be the way that we ordinarily speak. When we ask how so-and-so acted, we expect to be provided with a description of

his behavior alone. We assume that he intended this behavior, but we aren't looking for a description of his intention. (On the other hand, when we say that someone acted in a morally good or bad manner we *do* have mind some description of his intention. I will address this point in the second chapter.) When I speak of an act, then, I will mean someone's behavior that has been caused by his present intention to behave in that way. In abbreviated form, the stipulation is that *an act is intention-caused behavior*.

What is it, then, for someone to intend to behave in a certain way? In what does an intention consist? If an intention is a state, we can say that someone's intending to behave in a certain way consists in his having a "pro-attitude" toward his behaving in that way that involves his being disposed to behave in that way. Perhaps we can call this a "dispositional pro-attitude". If intention is a continuing event, we can say that someone's intending to behave in a certain way consists in his having a "motivating pro-attitude" toward his behaving in that way. Since I am already saying that intentions, rather than the onset of intentions, cause behavior, while inviting the reader to regard this as nothing more than a convenient abbreviation, I might as well call intentions "motivating pro-attitudes", while inviting the reader to similarly substitute "dispositional" for "motivating" if he is so inclined.

It seems that intentions aren't the only sort of motivating pro-attitude. There is also the pro-attitude that we ordinarily call "desire". I am about to propose that our desires only indirectly motivate us, however; they do so by causing us to have certain intentions. That there is a difference between intention and desire is shown by our complaints that we don't have any desire to do what we intend to do. A clue to what this difference might consist in is that when we say that we desire something, we think that

our having it would be pleasant for us, but this is not necessarily so with regard to our doing what we say we intend to do. A desire for something seems to be a motivating pro-attitude toward it, where it is the ultimate object of this pro-attitude. (The desired thing might be as modest as one's sitting comfortably at home or as grandiose as one's domination of the world.) This is perhaps why we think that our having what we want would be pleasant – pleasure seems to be linked to, if it is not identical with, the feeling that one has what one wants to have for its own sake. (I will use “wanting” and “desiring interchangeably.)

Someone might urge that someone's having an intention to do something is only his having a prevailing desire to get something in a certain way, together with a belief that he can get it in that way. I reject this proposal because it seems that we can both intend to do something that we don't want to do, and have an overwhelmingly strong desire to do something that we don't, and even *can't*, intend to do. Someone who urges that intending is only a prevalent wanting might assert that when we say that we intend to do what we don't want to do we are only distinguishing between desires for pleasant things and desires for things which can be had only at the cost of unpleasantness, and reserving the words “want” or “desire” for the former sort of desire. However, our apparent ability to want to do what we can't intend to do supports a literal reading of the ordinary usage according to which what isn't desired can be intended. A prisoner can have an especially strong desire to leave his prison, but he can't intend to do so unless he believes that he can pick the locks or tunnel through the walls. So, his especially strong desire to leave his prison can't in itself constitute an intention to do so. Someone who urges that intending is only a prevalent wanting might now assert that in such a case the

prisoner is only wishing, which is something different from desiring. However, when I imagine myself in the prisoner's position it seems to me that this "wish" would be the same mental state, undeniably one of desire, that later causes him to intend to leave the prison when he is informed that all of the locks have been unlocked and that all of the wardens are on strike.

I propose the following crude, but I hope useful, model of the relationship between intention and desire. *Our desires motivate us by together causing in us the onset of an intention to do what we think will bring about the satisfaction of one or more of them to a certain extent, during a certain time-period. We can have several desires simultaneously.* What sort of intention our desires will initiate in us depends upon factors including the comparative strengths of these desires, and our beliefs regarding how quickly and completely they can be satisfied in various ways (maybe one can be satisfied now, while another can only be partly satisfied in the distant future), and the degree to which a desire's satisfaction requires foregoing the satisfaction of other desires. The formula predicting what intention will be produced by certain combinations of desires, given such beliefs, must be very complex, and since it would in case vary from person to person (some people being better able to "defer gratification" than others, for example) it would be pointless for me to attempt to articulate it here.

We can imagine someone's various desires as vectors of force pointing in various directions. The sum of all of these vectors of force, as they are added to and subtracted from one another, can be called his "inclination." Relying upon this notion of "inclination" as a sort of averaging-out of desire, I will refer to the complex process by which someone's wanting various things causes him to intend to behave in a certain way

(given his beliefs about how he can get what he wants), by saying that someone's inclination causes him to intend to do what he is inclined to do.

Here it should be noted that someone may not want to do what he is inclined to do. He may be inclined to do something that he wants *not* to do, because he thinks that doing it will help him to get something that he wants, or that it will help him to partially satisfy more than one of his desires. I acknowledge that in ordinary speech we often use "inclination" to suggest an obvious desire for something; however, I think that in ordinary speech we may also use the word in the way that I use it. Thus, someone might say that he is inclined to tell the truth, and face the consequences, even though he doesn't want to, or that he is inclined to climb a steep hill in order to get to the party on top, even though he doesn't want to climb the hill.

It will be crucial to the coming discussion of the origination of intentions and acts that if all other relevant factors are fixed, then the stronger someone's desire for something becomes in comparison to his other present desires, the more the intention that his desires together initiate will be focused upon the satisfaction of that desire as opposed to the others. I will propose that someone can make himself intend to do what he thinks right, when he doesn't want to do it, by strengthening his desire to do whatever is right.

We can distinguish between more and less basic desires without abandoning the key point that the object of a desire, no matter how non-basic the desire may be, is still the *ultimate* object of the motivating pro-attitude. Thus, I may have basic desires for health and pleasure (I doubt that this is so, but let us suppose that it is.) Here I will assume that none of my other basic desires play any role in the production of my intention. So, if I only had my basic desire for health, and didn't have my basic desire for

pleasure, my desire for health would cause me to intend to eat something very healthy but unpleasant. If I only had my basic desire for pleasure, but didn't have my basic desire for health, my desire for pleasure would cause me to intend to eat something very pleasant but unhealthy. These two desires somehow combine to produce in me a less basic desire to eat something that is both somewhat healthy and somewhat pleasant, which in turn causes me to intend to find, purchase, and consume such a thing.

Someone's ultimate goal, in doing what he is inclined to do, might be something that isn't exactly any of the things that he wants. It might be something that only partly satisfies two or more of his stronger desires, by somewhat resembling two or more of the things that he wants. His having this as his ultimate goal would be a compromise solution, given his various conflicting desires. His *inclination* is then to do what he thinks will bring him to his ultimate goal. So, in the above example, someone's inclination is to buy some food that is somewhat healthy and pleasant, because the eating of such food is his ultimate goal. He doesn't want such food, though; he only wants healthy food, on the one hand, and pleasant food, on the other. Or, someone's ultimate goal might be something that he wants less intensely than he wants other things, because his stronger desires so directly conflict with one another that they cancel each other out.

In cases where someone's desire for X is so prevalent, albeit not altogether unopposed, that he is caused to do something *close* to whatever he thinks will get him X most efficiently, we would normally just say that someone's prevalent desire for X causes him to do the thing, Y, that comes close to being what he thinks will get him X most efficiently while also reflecting the slight influence of the opposing desires. We would normally omit any reference to an intervening intention to do Y. Still, though, as long as

the slight opposition of other desires causes the slightest deviation from the path that would get him X most efficiently, it seems necessary to distinguish intention from desire, at least as a sort of averaging of the various vectors of opposing desires. If intention is nothing more than an averaging-out of desires, then in cases where someone's desire for X is altogether unopposed this desire would simply constitute the intention to do Y, which is exactly what he thinks will get him X most efficiently. It seems to me, though, that even in such a case someone's intention to do Y must still be distinguished from the prevalence of his desire for X. If so, then intention is not simply an averaging-out of desires. However, here I rely on introspection alone, and admit uncertainty. I am even more uncertain with regard to cases in which someone's unopposed desire is simply to perform some core behavior, such as jumping or shouting, so that he would get what he wants by doing it. Here we don't even have a distinction between means and ends that would justify our distinguishing a means-oriented intention from an end-oriented desire.

Someone's ultimate goal in dancing, or in taking a walk, might simply be to dance or take a walk. He is therefore said to dance or take a walk "for its own sake." A murderer probably doesn't kill someone just for the sake of killing someone, though; at a minimum, he kills for the sake of (in order to bring about) his victim's death. The victim's death is his ultimate goal. Reverting, for the moment, to the notion of a "core" act as consisting solely in mental and bodily events, we can say, for example, that the murderer's overwhelmingly prevalent desire for someone else's death causes him to intend to extend his arm in a certain direction, etc., with the likely consequence that this person will die; the death of this person is then his ultimate goal in acting as he does. The dancer or walker's ultimate goal in performing his core act might be described as the

having of certain experiences – of rhythmic motion, in the case of the dancer, and of the traversing of a landscape as a result of muscular effort (the core act), in the case of the walker.

Someone's desires, like the intentions that they initiate in him, would be consciously "felt" by him. They would have to be, if they are pro-attitudes. We can take account of what people seem to refer to when they speak of enduring, long-term "desires" that persist even when the person that has them is thinking about things other than their objects by understanding these supposed desires as propensities to begin desiring the things in question. Napoleon doesn't desire mastery of Europe while he is flirting with Josephine, but he has a continuing propensity to frequently desire mastery of Europe. We can call the system of someone's propensities to begin desiring certain things his "personality". (If desires are states rather than continuing events, then they are or involve propensities to have certain intentions, while if intentions are likewise states they would in turn be or involve propensities to behave in certain ways. If we then introduce the proposed notion of personality, as seems necessary, we find ourselves speaking of propensities toward propensities toward propensities to behave in a certain way! This consideration supports the view that desires and intentions are continuing events rather than states.)

We intend to do something either "for its own sake," in which case we also want to do it, or we intend to do it for the sake of something else that we want. But what is to do A "for the sake of" B (where A and B may or may not be identical)? When someone does A for the sake of B his desire for B causes him to intend to A, but this is not an answer to the question. Perhaps the answer is that to do A for the sake of B is to intend to

try to get B to by doing A. “Getting B by doing A” is then the object of the intention, more fully stated. That this is so is suggested by the apparent interchangeability of “doing A for the sake of B” and “doing A *in order to* get B.” If someone asks me why I intend to do something that I don’t want to do, I might explain that I intend to do it in order to get something that I do want. This is pretty clearly the statement that I intend to strive for the thing that I want by doing the thing that I don’t want to do. (Here it might be objected that I would speak redundantly in saying that I intend to do A in order to get B, or for the sake of getting B, since in saying that I presently *do* A in order to get B I am already saying that my intention, in doing A, is to try to get B by doing A. But language is messy. I might also tell someone on Thursday that I intend to clean my apartment tonight in order to have a more enjoyable weekend.) If states of intending to do something (or continuing of events of intending, if that is what they are) can be picked out, so that the content of what is intended can be described – and this seems to be so, since we ask each other about our intentions, and offer answers to these questions – then it is surely the case that among these intentions are many whose objects are the doing of A in order to get B, or for the sake of getting B. We have in mind the image of somehow *aiming* our doing of A at the getting of B.

I take it that we never, or at rate almost never, perform a “core” act for its own sake. Even when I merely flex my muscles, I am doing so for the sake of a sensation, or for the sake of feeling that I have a good muscle tone, for example. Or, when I intentionally think about something, I do so not simply in order to think about the subject, but in order to gain some insight into it.

The claim that we can intend to do something in order to do, or for the sake of doing, whatever is right will be crucial to my account of what I will call “pure” moral responsibility. Among the things that we can want is our doing of whatever is right, and although we may not want to do something that we think it is right that we do, we may still intend to do it in order to do whatever is right. Our wanting to do whatever is right might cause us to intend this, if this desire of ours to do whatever is right is strong enough.

Note that talk of something’s *having* a desire or intention that it behave in a certain way can only be sensibly understood as a slightly abbreviated way of communicating the thought that it *desires* or intends that it behave in that way. For, after all, it isn’t as though “desires” and “intentions” are entities in their own right, that inhabit the people who “have” them. We are especially apt to begin thinking that in desiring things we *have* desires that inhabit us like parasites or, alternately, dwell outside of us, rubbing and jostling against us. These pictures lead people who speak of “desires” as causing behavior to suppose that people can be *forced* by their desires to behave as they do, as though these desires had seized control of their minds from within or were pushing them around like winds or large, unruly dogs. Perhaps the reason that we are not likely to think in this way of our intentions is that our intending to do something entails our thinking that *we* intend to do it, while in our desiring we may focus on the desired thing to such an extent that we forget about ourselves. In addition, we cannot have opposing intentions at one and the same time – although we can certainly have multiple intentions. Since the intentions that we have at any one time are mutually compatible, it is easier for us to feel “one” with all of them.

The proposal that our sympathetic, creative desires are internal to us, as states of our own being, while our selfish, destructive ones are distinct from us, merely inhabiting or surrounding us, is somewhat attractive. However, the proposal that *all* of our desires are like this is unacceptable. When we love someone, in directing our attention toward those of his specific affections and interests that seem good to us – that is, toward his apparently good specific desires (for example, my son’s desires for affectionate roughhousing and the weaving of fantasies involving talking machines) – we are sure that we are directing our attention toward *him*. We would reject any suggestion that we are only noticing alien entities, no matter how pretty these might be. We want to think that the specific desires toward which we direct our attention in loving him (and which we take to be the ultimate object of our attention) belong to the real inner core of his being. Even if these desires are expressions of a “deeper”, more general desire (a basic wanting) for the Good itself, and it is this desire that truly constitutes a person, they are nevertheless continuous with this basic desire for the Good that they express, and are therefore internal to the person.

It is harder to imagine desires in general as external invading or obtruding entities when, rather than saying that someone’s *desires* cause him to behave in a certain way, we say that his *desiring* that he behave in a certain way causes him to behave in that way. Still, talk of someone’s desiring something suggests talk of his having a desire for it. It is safer to speak of someone’s “wanting” to behave in a certain way rather than of his desiring to behave in that way, since talk of someone’s “having wants” is not interchangeable with talk of his wanting things, as talk of his having desires is

interchangeable with talk of his desiring things; talk of “wants” suggests privation and need, rather than supreme desire.

Theorists who hold that the “self” is wholly constituted by a body’s mental states and dispositions toward being in certain mental states, and/or by mental events and objects (“ideas”), may prefer to speak of “desires” and “intentions” rather than of someone’s wanting and intending things, so as to eliminate the reference to a thing that would be distinguishable from its intending in the same way that a body is distinguishable from its falling or spinning. However, since my purpose is to sketch out a plausible picture of someone’s making himself intend something I am committed to the view that persons are subjects, not systems, of desires, intentions and beliefs. This view is most clearly communicated by speaking in the ordinary way of *someone’s* wanting something or intending to behave in a certain way, and of his wanting something causing his intending to behave in a certain way, and of his behaving in that way therefore being his acting in that way. Nevertheless, it will often be convenient to formulate this in abbreviated form by saying simply that acting is behaving caused by intending, or to drop this way of talking altogether and to say instead that desires cause intentions, which in turn cause behavior.

Carl Ginet attacks the notion that “motives” are force-endowed states whose relative strengths determine behavior, with the strongest motive prevailing in cases of simple opposition. What he refers to as “motives” seem to be exactly what I refer to, more colloquially, as “desires”. Ginet deploys Thomas Reid’s argument that without some “test” of a “motive’s” degree of forcefulness beyond the degree to which it prevails, talk of the motive’s degree of forcefulness either refers only to the motive’s degree of

prevalence or else assumes what is in question – that a motive is endowed with some power whose quantity causes it to prevail to that degree. Presumably, the “test” would consist in checking some introspectively evident indicator of a motive’s intensity. In any case, the forcefulness of a motive would have to manifest itself in some way other than through the occurrence of behavior that satisfies the desire. Ginnet rejects Mill’s suggestion that “the degree of pain and pleasure anticipated” is the requested criterion, pointing out that we occasionally select an action anticipated as less pleasant than the alternatives.⁴

Ginnet follows Mill in taking Reid’s discussion of “motives” to be a discussion of “desires”.⁵ Indeed Reid’s “motives” *do* seem to be desires, except that they include both the sorts of desires that we share with chimpanzees and uniquely human desires to do what “...is our duty, or conducive to our real good...”⁶ Reid calls the former desires “animal motives”, and the latter “rational motives”.⁷ In fact, Reid *does* think that both sorts of motives have quantifiable strengths, and that is precisely why it makes sense for him to call them “motives”; only, he doesn’t think that the strengths of animal desires are commensurate with the strengths of rational desires, or vice-versa. Rather, animal desires are only commensurate in strength with one another, and the same is true of rational desires. In “brute animals”, the strongest animal desire indeed “always prevails”, and there *is* a test of an animal desire’s degree of strength when it is *our* animal desire, from which it can be inferred that it is meaningful to speak as well of varyingly forceful animal

⁴ See *On Action* 131-136, referring to Thomas Reid, *Essays on the Active Powers of the Human Mind* (Cambridge: M.I.T. Press, 1969), 288-289, and John Stuart Mill, *On the Logic of the Moral Sciences: A System of Logic, Book VI*, ed. Henry M. Magid (Indianapolis: The Bobbs-Merrill Company, Inc., 1965), 160-161.

⁵ See Mill, 158-159.

⁶ Reid, 290.

⁷ Reid, 289.

desires in “brute animals”. The test of an animal desire’s degree of strength in us is the degree of effort required for our resistance to it.⁸ The test of a rational desire’s strength in comparison to that of other rational desires, however, is only this, that “the strongest, in the eye of reason, is that which it is most our duty and our real happiness to follow.”⁹ Since the strengths of our animal and rational desires are incommensurable, so that “[a] motive which is the strongest, according to the animal test, may be, and very often is, the weakest according to the rational,”¹⁰ our acts cannot be determined by a simple playing-out of competing forces.

Reid does not deny that desires are forceful; rather, he only denies that the forcefulness of animal desires is commensurable with the forcefulness of rational desires. Ginet, however, denies that desires have any sort of behavior-causing force. It seems to me, though, that if we deny that desires are forceful then we have to deny that they are anything at all, beyond collections of mental images, fragmentary mental verbiage, and real and imagined physical sensations and twinges. If this is all they are, then they can play no role in the explanation of our behavior. But they evidently *do* play a role in the explanation of our behavior, and the role they play is that of a disposition, like the role of the weight of a body in the explanation of the body’s descent, or even, if they are continuing events, that of efficient cause. Moreover, we quite frequently say that we want something more now, or less, than we did yesterday, or that we want this more than that. When we sincerely assert that we want to do something more now than we did yesterday, we are more likely to do it today than we were yesterday. When we sincerely assert that we want this more than we want that, it is likely that we will do the thing that we say we

⁸ Ibid., 289.

⁹ Ibid., 290.

¹⁰ Ibid., 291.

want more. A reasonable explanation of this is that in asserting that we have a greater quantity of desire we are asserting that our desire is more forceful. Furthermore, the fact that we try to excuse ourselves from moral responsibility by imagining that we are driven into some bad behavior by desires sitting within us like parasites or carrying us along like wild horses or shoving us around like unruly dogs is itself evidence for the forcefulness of our desiring, even though this image is a wish-fulfilling misrepresentation. We wouldn't be able to deceive ourselves in this way about desires if their real forcefulness didn't make this possible.

Alfred Mele argues on conceptual rather than experiential grounds for the meaningfulness of regarding desires as forceful, or as having “motivational strength”. He argues that the characterization of behavior as intentional presumes that the behavior is the object of a motivating “want or desire”, and that “once we grant motivation or wanting a genuine place in our conceptual scheme, motivational *strength* gains a foothold as well.”¹¹ Mele's argument depends upon the identification of desire with motivation – by which he seems to mean something like “behavior-causing power” – and so might be thought to beg the question, assuming from the outset the conclusion that desire is something that should be understood as a force. This assumption is only justified if it fairly represents our experience.

I don't regard the forcefulness of desires as immediately directed toward the causing of the desired behavior, though. Rather, I regard it as immediately directed toward the causing of intentions to behave in the desired way. If someone were really “overwhelmed by passion”, seeing red and striking out in a frenzy, or “succumbing to

¹¹ *The Springs of Action: Understanding Intentional Behavior* (Oxford University Press: New York 1992), 82-83).

lust”, without having an intention to strike or embrace, he would be mentally impaired – a suitable target for medical intervention. It should also be remembered that I regard desires as necessarily “conscious” – that is, that I regard desiring as a state of consciousness. What we might otherwise think of long-term, standing desires that repeatedly rise into and then subside out of consciousness, so that sometimes we think about the thing that we desire and sometimes don’t, would instead be propensities to desire certain things with certain degrees of forcefulness in certain situations. The overall system of such dispositions would be what we call someone’s “personality.

Given my stipulated use of “action” to designate behavior that is caused by intending, I have to say that intending to *act* in a certain way entails but is not limited to intending to behave in a certain way. Someone’s intending to act in a certain way would be his intending to behave in a certain way that he thinks of as being caused by his intending of it. It seems, then, that we rarely, if ever, intend to act. But can we *want* to act in a certain way, given that this would mean our wanting to intend to behave in that way?¹² Given my account of desires as *motivating* pro-attitudes, we could want to intend to behave in a certain way only if this wanting could cause us to intend to do something in order to bring about our intention to behave in that way, or at least in order to make it more likely that we will have this intention. This seems possible. I might want to intend, tomorrow morning, to study. This desire might cause me to intend to clean off my desk tonight, so that I will have this intention tomorrow when I see it. Or, suppose that a drug

¹² Our wanting to intend to behave in a certain way suggests Harry Frankfurt’s picture of the “freely acting” person as wanting the desire that causes his behavior (his “will”) to cause his behavior. (See “Freedom of the Will and the Concept of a Person” [1971] in Laura Waddell Ekstrom, ed., *Agency and Responsibility: Essays on the Metaphysics of Freedom* [Boulder: Westview Press, 2001], 77-91.) However, to intend to do something cannot simply be to have what Frankfurt calls an “effective” desire to do it – a desire that causes us to do it if nothing stops us from doing it (80). A prisoner might have an unsatisfiable effective desire to leave his cell – effective because it would cause him to leave his cell if its door weren’t locked – but cannot intend to do so, give his knowledge that the door is locked.

addict wants to stop intending to drug himself. It wouldn't be enough, given that desire motivates, to say that the addict's thought of his recurrent intention to drug himself is painful to him. His desire not to have this intention would have to be capable of making himself intend to do something to bring about his no longer so intending. Well, perhaps he thinks that if he were to get regular exercise he would no longer intend to drug himself. So, his desire not to intend to drug himself causes him to intend to get regular exercise in order to remove the intention to drug himself.

If desire and intention are what I say they are, then what is it for someone to choose and decide? This is what I propose: *Someone's "choosing" to do something is his beginning to intend to do it when he recognizes that he might do other things instead. His "deciding" to do something is his choosing to do it after deliberating over these various possibilities.*

The account of intentional behavior that I advocate is more complex than one according to which someone's doing something intentionally would simply be a matter of his having a dominant desire to do it, and believing he can do it.¹³ I hold that his desires *cause* him to intend to do something, so that his intending is distinct from his desiring. I hold this because our ordinary inclination to describe someone as intending to do something that he doesn't want to do seems to reflect our experience of different pro-attitudes which we call by these different names, "intention" and "wanting".¹⁴ But since I still see someone's intentions as normally determined by the makeup of his desires, and as causing his behavior in their stead, I am closer to this approach than to an alternate one

¹³ Perhaps Frankfurt would explain intentional behavior in this way.

¹⁴ See Lawrence H. Davis, *Theory of Action* (Prentice-Hall: Englewood Cliffs, 1978), 7. Davis formulates the proposal that he is targeting as one according to which "an action is a doing in which a doing-related event occurs because the agent wants it to, then and there" (5).

according to which what distinguishes what we would ordinarily call “intentional” behavior from “unintentional” behavior is causation by a special sort of mental event, a “volition.” Advocates of a volition-based approach typically call volitions “acts” in their own right.¹⁵ Presumably, though, they don’t mean to claim that these would be instances of intentional behavior – as though someone must intend to have a volition if he is to have it. Rather, these are instances of mental behavior that is distinguished from other mental behavior that isn’t action (such as daydreaming or having a panic-attack) by what Carl Ginet calls its “actish phenomenal quality”. It would be most evident to us that mental events with this quality are acts in their own right when they fail to cause the “willed” behavior; otherwise, our experience of them would tend to be absorbed into our experience of the willed behavior. However, they would be acts in their own right whether or not they cause the willed behavior. Given that there is a sort of mental event distinguished by this actish phenomenal quality, the objection that the volition-based approach relies upon purely hypothetical, never-experienced events fails.¹⁶ However, it seems to me that if volitions are indeed experienced in this way, then a volition is simply the onset of an intention; to have a volition is to begin to intend.

Defenders of a volition-based model of action might be motivated to distinguish their volitions from the beginnings of intentions by the observation that there are various

¹⁵ Carl Ginet proposes that what makes behavior an act is that it is either a “volition” or consists in the combination of a volition with behavior caused by this volition (*On Action* [Cambridge: Cambridge University Press 1990], 15). Ginet thus pictures “volitions” as actions in their own right that are internal to complex actions combining volitions with the behavior that they cause. While this proposal is further-elaborated in the following pages, it remains the core of his account. A similar position is endorsed by Lawrence H. Davis (*Theory of Action*, 15-26). In the 18th century, Thomas Reid made “determinations of the will” requirements for “voluntary” action, proposing that “[I]n every voluntary action the determination of the will is the first part of the action...” (*Essays on the Active Powers of the Human Mind*, 264). Here “voluntary” seems redundant. The question of whether we call “volitions” acts in their own right, or components internal to minimal acts that are conceived of as complex events, or as events that aren’t acts but that in causing behavior qualify that behavior as action seems purely verbal.

¹⁶ Richard Taylor so objects, in *Action and Purpose* (Englewood Cliffs: Prentice-Hall, Inc. 1966), 66-69.

kinds of cases of ostensibly intentional behavior in which there is no corresponding “intention” – no relevant pro-attitude – seems to be operating. Thus, Lawrence H. Davis offers as examples a driver “stepping on the brakes” to avoid colliding with someone, someone doing something “absent-mindedly”, and someone impulsively expressing strong feelings.¹⁷ Davis’s denial that there is a relevant intention in the first and last cases depends upon the premise that an intention to behave in a certain way must entirely precede the behavior, however, so that “[i]f Sam intends to do an A, he has not yet done it.”¹⁸ This seems to me to contradict our ordinary experience and way of talking about intentions. I ask someone who is annoying me, “Are you *intending* to annoy me?” My intention to go to Manhattan continues throughout my ride on the R train, which constitutes my going to Manhattan, the intended event. As for the case of someone doing something absent-mindedly when he intends to do something else, it seems to me that his intentional act would consist in the portion of what he is actually doing that he intends to do. If, intending to comb his hair with a comb, he absent-mindedly combs his hair with a fork, his intentional behavior consists in running a hard, pronged object through his hair. If no intention at all is (or was) present, then he doesn’t engage in any intentional behavior at all – he doesn’t act.

Richard Taylor offers a third sort of account of what distinguishes the intentional behavior that I call “action” from behavior that I would call “mere” behavior. According to Taylor, someone’s action is behavior (his own, of course) that is caused by *him*, rather

¹⁷ *Theory of Action*, 7. Davis here formulates the proposal that he is targeting as one according to which “an action is a doing in which a doing-related event occurs because the agent wants it to, then and there” (5), but then he later revisits these examples to show that not all intentional action involves an intention to behave in the way in question (59-60).

¹⁸ *Ibid.*, 59.

than by his wanting or intending it, or by his volition.¹⁹ Taylor holds that someone might either be the first cause of his behavior, or he might be caused, by something other than himself, to cause his behavior. I don't see, though, how someone could be caused by something other than himself to cause his behavior except by being caused to intend this behavior. But in this case, what would make his behavior action would be that it is caused by his intending it. This is the proposal that I endorse. Taylor's proposal then becomes the proposal that behavior is action just in case (a) it is caused by the behaving thing's intending it (the proposal that I endorse), or (b) the behaving thing is its first cause. But the causing of behavior by the behaving thing's intending it seems too different from the "first-causing" of behavior by the behaving thing to warrant the application of the same name, "action", to the behavior in both instances. This is one reason to reject Taylor's proposal.

Even if Taylor thinks that someone can be caused to cause his behavior in some other way than by being caused to intend this behavior, surely Taylor would have to allow that someone can be caused to cause his behavior only by being made to have some sort of strong propensity to cause it. But then why can't we say that inanimate things can be made to cause *their* behavior by being made to have some sort of strong propensity to cause it? For example, take Taylor's own comparison of the starting a forest fire by a match to the starting of a forest fire by a man.²⁰ The ignition of the match might be understood as the striking-event's making the match cause its (behavior of) catching fire, just as it might be thought that someone's having been fired, or left by his spouse, causes

¹⁹ See *Action and Purpose*, 111-112 ("Agents as Causes"). Taylor regards "volitions" as unnecessary hypothetical entities (66-69), and doubts that anyone's desire to behave in a certain way (his "wanting" to behave in that way) can cause him to behave in that way.

²⁰ See *Action and Purpose*, 120-122.

him to cause his subsequent drinking binge. True, this analysis of a match's ignition is not the usual analysis in terms of a causally linked series of events, in which substances play a role only as locations for these events, that is championed by Taylor himself where the behavior of things like matches rather than men is concerned.²¹ It has, however, received some recent support.²² If we accept this sort of picture of causation involving inanimate things, then we cannot simply distinguish acts from events that aren't acts on the basis of the causation of the former by substances, as opposed to the causation of the latter by other events. It turns out that *all* causation is by substances. It must be *how* a substance causes its behavior that makes the difference, then. The most plausible proposal regarding the difference between how substances cause the events that we call acts and how they cause the events that we don't call acts seems to me to be that they cause the events that we call acts by intending to behave in those ways, and that they cause the events that we don't call acts in some other way. But this is the first formulation of what action is – the formulation that I endorse.

Roderick Chisholm agrees with Taylor in holding that an act is caused by the behaving thing, without any intermediation of mental events or states such as volitions and desires (or intentions) as proximate causes. However, he escapes the collapse of this formulation into the one that I endorse, according to which an act is behavior caused by an agent's intending it, because he doesn't entertain any notion of an agent's being caused to cause his behavior.²³ However, Chisholm thereby makes it impossible for us to

²¹ “[The match] only, unlike the man, underwent a change, induced by something other than itself, which was causally significant in that series of changes culminating in the forest fire” (Ibid., 122).

²² See Timothy O'Connor, *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press, 2000), 68-71.

²³ See “Human Freedom and the Self” [1964], in Ekstrom 2001, 126-137. (There *are* intermediate physical events or states, though, according to Chisholm, at least where physical behavior is concerned; someone

distinguish between cases in which people are morally responsible for their acts and cases in which they aren't. The significance of this objection to Chisholm's proposal can only emerge later in this discussion, when we consider what moral responsibility is, and what it requires.

If we wish to preserve the thought that something can be a first cause of its action, we can do so without requiring that the name "action" be applied only to behavior that the behaving thing causes. Following my stipulation that the name "action" be applied to behavior caused by the behaving thing's intending to behave in that way, we can hold that something can be a first cause of its intending to behave in a certain way. Or, we can apply the name "action" to the combination of something's intending to behave in a certain way with its behaving in that way, when its intending to behave in that way causes its behaving in that way, and hold, again, that something can be a first cause of its intending to behave in a certain way.

The idea that people, substances, can be causes is offered as an analysis of the image that we sometimes have of ourselves as being the sources, or original producers – the "originators" – of our acts. I will propose an "event causation" analysis of the same image that I think does just as much justice to it.

causes the motion of his arm by causing an event in his brain.) But unlike Taylor, Chisholm doesn't entertain the notion that an agent might act unfreely, by being caused to cause his behavior.

2. *Freedom and moral responsibility*

The following definition of moral responsibility seems both to fit our intuitive sense of what moral responsibility is and to account for our attribution of moral responsibility to some kinds of things and not to others. *Person P is morally responsible for X just in case everything is true of P that must be reasonably believed about P both by anyone who in considering X should like P as a person and by anyone who in considering X should dislike P as a person.* Note that it is far from sufficient simply to say that someone is morally responsible for his intention just in case anyone should like or dislike him as a person while considering it. Anyone should like or dislike him as a person in considering his intention while reasonably believing certain things about him. However, these beliefs, though reasonable, might be false. If they are false, then although one should dislike or dislike him as a person while considering his intention, he is not morally responsible for it.

This definition of moral responsibility depends upon the possibility that one person should like someone as a person (with respect to his personhood) in considering his act, while another person should at the same time dislike him as a person in considering his act, because they have differing, although reasonable, beliefs about him over and above the reasonable beliefs that they share about him.

In saying that someone “reasonably” believes something, I rely upon our ordinary use of this word, and upon the ordinary judgment that it expresses. I won’t venture to define “reasonableness”. Someone’s ability to give reasons for his belief that have some persuasive force even for impartial people, and his believing what he believes because of the same experiences and inferences that he alludes to in giving these reasons, surely play

a large role in making him reasonable. We ordinarily distinguish reasonable beliefs from unreasonable ones without having any precise formulation of what reasonableness is ready to present upon demand.

By “in considering X”, I mean “while and as a result of considering X.”

In saying that someone *should* like or dislike another, I mean that this is the right attitude (although not necessary the best one with respect to its overall, longterm effects upon the world) for him to take toward the other. I don't take a position on what rightness and goodness are, and on how they are related. I rely upon what I take to be the reader's intuitive recognition of the rightness and goodness of things. In ordinary usage, several alternatives can be right without any being *the* right one, and anything that isn't right is to a greater or lesser extent wrong. I will stretch ordinary usage by calling things more or less *right*, insofar as they are less wrong. A partial fulfillment of a promise, for example, is less wrong, and thus more right, than complete non-fulfillment of it.

A “person” is at the very least a self-conscious, desire-endowed, rational agent. A rational agent is one able to conceive of a wide range of possible acts, and to formulate complex strategies for attaining ultimate ends. Since animals and very small children are not rational agents, they are ruled out from the start as candidates for moral responsibility. No one should like or dislike them as persons, since persons must be rational agents, and not being rational agents they are not persons. This is entirely compatible with its being right that one sometimes like or even dislike them in considering their acts. It just isn't ever the case that one should like or dislike them *as* persons in considering their acts. In using the expression “as a person” I mean not only to distinguish liking or disliking someone as a person from liking or disliking him as, say,

an athlete or pianist, but also to convey that liking or disliking him does not extend beyond his bare personhood to, say, his being one's friend or teacher. In liking one's friend as one's friend, one certainly likes him as the *person* who is one's friend. Perhaps, though, one should like one's friend not merely as a person, but also as the person who is one *friend*, in considering his act. Perhaps, then, one should like him to a greater extent in considering his act than one otherwise would. I want to convey that the liking relevant to moral responsibility is impartial. It is the liking that remains when the thought of someone's special role as one's friend (or whatever) is removed.

Small children presumably acquire more and more moral responsibility for more and more of their acts as they develop into persons. Similarly, psychotics would lose and regain moral responsibility by degrees as they descend into and emerge from their psychosis. Personhood is not all-or-nothing. For this reason, "allowances" are made even for the wrong acts of sane adults, when they act in conditions of stress that seem to reduce the degree to which they can act as persons.

I speak of liking and disliking rather than of "pro" and "con" or "positive" and "negative" attitudes toward someone because liking and disliking have the same generality while retaining the emotional resonance of ordinary speech. Liking or disliking someone as a person when it is possible to become or remain personally and/or communally involved with him seems to entail a positive or negative attitude toward such involvement with him.²⁴ A positive or negative attitude toward communal involvement

²⁴ Here I have in mind Peter Strawson's conclusions regarding the link between attributing moral responsibility for his acts to someone and assuming "participant reactive attitudes" toward him – attitudes that we take with someone with whom we *participate*, or are ready to participate, in an "interpersonal relationship" ("Freedom and Resentment" [1962], in Ekstrom 2001, 183-204; 190,192). However, Strawson sees what he calls a "negative" participant reactive attitude toward another as entailing an acceptance of full involvement with him rather as a repudiation of involvement with him or a readiness to

with someone would consist in a readiness or reluctance to have someone as a member of one's real or imagined community, and to act toward and together with him accordingly.

This would entail a readiness to have him fully participate in whatever activities are central to the idea that one entertains of one's community – from productive work to ritualized worship and celebration, in which the “we” that works, worships and celebrates is felt to include him as a full-fledged participant. Perhaps a readiness for communal involvement with someone is a precondition for a readiness to become personally involved with him as well, whether in a closely collaborative relationship or in a friendship, or in both.

Liking someone as a person when one is not presently involved with him seems to entail a readiness to become involved with him; disliking someone as a person, when one is not presently involved with him, seems to entail a rejection of the possibility of such an involvement. Liking someone as a person when one *is* presently involved with him seems to entail a readiness to sustain and perhaps deepen the involvement; disliking someone as a person when one is presently involved with him seems to entail a readiness to withdraw. Liking or disliking someone as a person when he is dead or too distant for there to be any likelihood of involvement with him might entail a sort of imaginary readiness or reluctance to enter into or sustain an involvement with him. However, liking and disliking someone as a person may entail more than these positive or negative attitudes toward a possible or imagined involvement with him. Because more than this may be entailed, I confine myself to speaking generally of liking or disliking someone as a person rather than more specifically of having a positive or negative attitude toward

withdraw from such an involvement. Further, Strawson doesn't say that we ever *should* participate in an interpersonal relation with someone. Finally, Strawson focuses exclusively upon one-on-one, or “interpersonal,” as opposed to communal, involvements with people.

involvement with him.

It seems that people should like or dislike someone as a person in considering X only if they reasonably believe X is his intention or act, or some consequence of his act, or an event that would not have occurred had he acted differently. So, someone can be morally responsible only for his intention or act, or some consequence of his act, or an event that would not have occurred had he acted differently. It may be objected that we ordinarily say that someone is morally responsible for his *failure* to do or to intend to do something that he might have done or intended to do, or for a consequence of his *inaction*. Thus, we might explain the notion that a parent is morally responsible for the wellbeing of his children by saying if his children come to harm as a consequence of his inaction, he will be morally responsible for this consequence. However, it is simpler to think of a failure to do or to intend to do A as an intention to do B, when A might have been done or intended. We can also avoid speaking of someone as being morally responsible for a consequence of his inaction by saying that he is morally responsible for an event that would not have occurred had he acted differently.

It seems clear that whenever people should like or dislike someone as a person in considering his act or some consequence of it, they must consider his act or its consequence as the consequence of his intention, and so would be considering his intention. Whenever they should like or dislike someone as a person in considering an event that would not have occurred had he acted differently, they must consider the act that he *did* perform as the consequence of his intention, and so, again, would be considering his intention. Their reasonable belief regarding the character of his intention is always what makes it right that they like or dislike him as a person. So, whenever

people should like or dislike someone in considering his intention, act, or an event that would not have occurred had he acted differently, they should like or dislike him in considering his intention alone.

A person “credits” someone for his intention (or “gives him credit” for it) in liking him as person while and as a result of considering his intention. A person “blames” someone for his intention in disliking him as a person while and as a result of considering his intention. The things that the people who should blame someone for his intention *and* the people who should credit him for it must reasonably believe about him are the “requirements for moral responsibility.” If these things are true of him, he “meets the requirements for moral responsibility.” If all of the things that anyone must reasonably believe about someone whenever he should give him credit for his intention are true, then the person who should be given credit *deserves* this credit. If all of the things that anyone must believe about someone whenever he should blame him for his attention are true, then the person who should be blamed *deserves* this blame.

Note that, given the definition of moral responsibility offered at the beginning of this chapter, someone can be morally responsible for an intention for which he deserves neither credit nor blame. As long as everything is true of him that must be reasonably believed about him both by anyone who should give him credit for his intention and by anyone who should blame him for it, he is morally responsible for it. It might be right for someone to give him credit or blame for his intention even though he doesn’t deserve this credit or blame, since the reasonable beliefs that make this right might be false.

Accounts of what moral responsibility requires are sometimes offered as accounts of what acting freely requires, or assume that the requirements for performing an act are

also the requirements for being morally responsible for it. Such accounts propose requirements for what they call “acting freely” that indeed reflect what we often have in mind when we say that someone has acted freely or unfreely. It seems to me, however, that we often say this only because we mistakenly imagine him as dissociated from his desires, or from those that are bad, or irrational, or unwanted, and so forth. When we think such desires play a predominant role in producing his intention, we think of them as forcing him to have this intention. We then think of him as acting unfreely. When we think that his intention has not resulted from their influence, or has resulted from the predominant influence of desires that we do not imaginatively dissociate from him, we think of him as having acted freely.

When we say that someone “acted freely,” we ordinarily mean that he acted without being forced to act as he did. If acting is behaving caused by intending, then someone would be forced to act as he does by being forced to intend to behave in a certain way. Something is forced to behave in a certain way only if it is caused to behave in a manner other than that in which it tends on its own to behave. It is forced to assume a certain state only if is caused to assume a state that is incompatible with the state that its own internal propensities would cause it to assume, given its circumstances. So, something’s own propensities cannot force it to behave in a certain way, or to assume a certain state. Thus, a rock’s own weight (a gravitational propensity) cannot force it to roll downhill. Similarly, someone’s own personality (a system of propensities to desire certain things with certain degrees of intensity under certain conditions) cannot force him to intend to behave in a certain way. Nor, really, can someone’s desiring something force

him to strive for it, no matter how intensely he may desire it, although we may imagine that he is forced because we dissociate his desires from him.

Sisyphus can force his rock to roll uphill. Similarly, a future neurologist might force someone to intend to behave in a certain way, and therefore to act in that way, by electrically stimulating his brain in a way that causes him to have desires that he has no propensity to have – desires that are contrary to his personality. Since these desires are contrary to his personality, the resulting intention and act are also contrary to his personality, and so he acts unfreely. The assumption here is that the patient has not chosen to have his brain stimulated. If the patient had chosen to have his brain stimulated, perhaps because he wanted to see what it would be like to have these strange desires, then the resulting act would express his personality in a round-about way, and so would be free. Even if the patient *would* have chosen to have his brain stimulated, but hasn't so chosen, his act is forced. It is contrary to his personality. Maybe his personality does include an odd propensity to want external things (such as neurosurgeons) to make him have desires that he wouldn't otherwise have. However, unless his desire for this manipulation itself plays a large role in causing it, the desires resulting from the manipulation don't express his personality.

A neurologist who only electrically stimulated someone's brain in a way that directly causes him to have certain thoughts and to move his body in certain ways would not be forcing him to *act* in that way, for he would not be forcing him to *intend* to behave in that way; he would only be forcing him to behave in that way.

Although we would say that Sisyphus forces his rock uphill, we would not say that in doing so he forces it to roll *downhill*. The reason that we would not say this is that

once the rock is on the hillside, and nothing prevents it from descending, its own weight causes its descent. Why, then, are we inclined to say that torture forces a patriot to surrender information damaging to his country? Here, the torture puts into effect his propensity to desire something with a certain degree of intensity (the cessation of pain) in certain general circumstances (extreme pain), and this causes him to form intentions that fit his new configuration of desires. Doesn't the rock's relocation up the hillside similarly put into its effect its propensity to move downward whenever possible? We do *not* say that the moving of the rock up the hillside forces it to roll downward, so why do we say that torture forces someone to act as he does?

Torture forces a patriot to talk only if talking is contrary to his personality. Personality is a system of propensities to have certain desires, and so is analogous to the rock's gravitational propensity. To solve this problem, then, we need to see how the patriot's divulgence of information is contrary to his personality. We can say, I think, that a patriot is forced to divulge information under torture – that the divulgence is unfree – if (a) because of his circumstances (being tortured), no other act seems possible to him, (b) he would choose not to be in these circumstances if possible, and (c) if he weren't in circumstances making no other act seem possible to him he would not divulge the information. Someone who leaps out of the window of a burning building in order to escape the flames is similarly forced to so, because he meets the same conditions. No other act is thinkable, he wouldn't be in the burning building if he could help it, and if he weren't in circumstances that make doing anything other than leaping out a window unthinkable he wouldn't leap out of the window. His act is contrary to his personality.

This way of understanding the tortured patriot's unfreedom, or that of the man in the burning building, requires that we see a drug addict as acting freely when he takes his drug in order to alleviate his withdrawal symptoms or to satisfy a craving that he wishes he didn't have. Similarly, it requires that we see someone as acting freely when he stays home so as not to get rained on. It must seem *possible* to the drug addict (unless his withdrawal symptoms are truly torturous) to not take the drug, or to the person who stays home to go ahead with his plans despite the rain. We do ordinarily regard the drug addict as taking his drug freely, and the person who stays home when it rains as doing so freely. This is not to say that we don't also think of the drug addict as a "slave" to his addiction, and of someone as being forced by the rain to stay home. Of course we do. But strict consistency with our ordinary judgments regarding the freedom and forcing of physical motion (as in the case of the rock) requires that we prefer the ordinary assessment of the matter according to which the addict and the rain-averse person act freely.

It might at first appear that my justification for regarding the tortured patriot as being forced to talk also justifies the conclusion that a man who assaults someone in a fit of rage does so freely. The enraged man sees no other act as possible, would not have chosen to be in the situation that he is in, and would not assault anyone if he weren't in such a situation. However, the tortured patriot and the enraged man are distinguished by the different reasons for their inability to see any other act as possible. The tortured man's *circumstances*, circumstances that he would not have chosen, prevent any other act from seeming possible to him. The enraged man's own *personality* prevents any other act from seeming possible to him. So, his act very much expresses his personality, and as such is free.

Perhaps someone might want to say that crimes of passion are *not* free acts. However, if they are not, then the passionate criminal doesn't deserve blame for his crime any more than someone would deserve blame for being shoved into someone else by a third party. If the passionate criminal is blameless, then he should not be punished. Instead, he should be treated in the way that psychotics are treated when they behave destructively. This would require such sweeping changes in our penal system that the burden of proof is surely on those who would assert that such crimes are not freely committed.

Alternately, someone might want to say that while the tortured patriot freely divulges his country's secrets, he isn't morally responsible for doing so. Perhaps those who make this claim would explain his lack of moral responsibility for what he does by saying that if his desire not to be in pain is so overwhelming that he really can't think of doing anything other than what would enable him to satisfy it, then in betraying his country he isn't acting as a *person*. He isn't acting as a person, because a person is a rational agent, and a rational agent is one able to conceive of a wide range of possible acts, and to formulate complex strategies for attaining ultimate ends. The urgency of his desire to escape pain prevents him from conceiving of any acts beyond the one that will bring an end to his agony. So, no one with an adequate view of his situation should like or dislike him as a *person* in considering his act. It seems to follow, then, that he is not morally responsible for it.

However, the same consideration would then have to excuse someone who kills in a fit of rage from moral responsibility for his act. Such a person often can't think of doing anything other than assaulting the object of his anger, just as the tortured patriot

can't think of doing anything other than what would bring an end to his agony. But if our present penal practices are right, then the passionate criminal must be morally responsible for his act. So, someone's loss of ability to conceive of alternate acts can't make him lose his moral responsibility. Surely, the tortured prisoner *isn't* morally responsible for his divulgence of information. Since his lack of moral responsibility can't be due to his inability to conceive of alternate acts, it must be due to his lack of freedom. He is unfree in that his inability to conceive of other acts is forced upon him. He isn't morally responsible for his act because he acts unfreely.

How, then, can we respond to the argument that since someone who acts under the influence of a desire so urgent that he cannot think of anything other than satisfying it does not act as a person, the enraged man cannot be morally responsible for his assault? I would admit that the enraged man does not act as a person. The extreme urgency of his desire to do harm to the other man breaks down his personhood. However, part of what leads us to like or dislike someone as a person is our sense of how resistant he is to stresses that will result in the breakdown of his personhood, and also our sense of the ways in which his personhood might break down: is he prone to fits of rage, panic, lust, or despair? We should dislike the violent man as a person *in* considering his act because we already know that he is a person; now we know as well that his personhood tends to break down in fits of rage when he is opposed. (Note that what someone does when his *personhood* breaks down might still express his *personality*.)

To say that someone is morally responsible for doing what he does when his personhood breaks down is not to say that he necessarily deserves blame for this act. One should not necessarily dislike him as a person in considering what he does when his

personhood breaks down. Whether or not one should dislike him depends upon what is normal (in the sense of meeting an ideal standard) where personal fortitude is concerned. One should dislike someone who is *abnormally* prone to breakdowns of personhood that lead him to perform a wrong act. Here the assumption is that the act is wrong. An act's rightness or wrongness is independent of the integrity or collapse of the agent's personhood. Let us suppose, for example, that assaulting someone except in cases of necessary self-defense is wrong. If a normal person would not have assaulted another in a fit of rage under a given set of circumstances, someone who does deserves blame.

Nothing forces a leopard to go hunting or prowling for a mate. Its desires for food and sex, which cause its behavior (perhaps via an intermediary intention, although this is doubtful), cannot force it to do so, because they are its own states (or ongoing events of which it is the subject). Similarly the powerful short-term desires that cause children to rapidly form and abandon intentions to do what they most want to do are these children's own states, and so cannot force them to do what they do. And the same is true of the man who guns down the businessmen whom he believes to be alien invaders. Here, perhaps, it is his wildly false beliefs, not the strengths of his basic desires, which make him crazy. Still, since he is the subject of these beliefs, they cannot force him to do what he does. But animals, small children, and psychotics cannot be morally responsible for their acts, because they don't act as rational agents, and therefore don't act as persons. So, freedom isn't sufficient for moral responsibility.

Many people say that someone must be able to *originate* his acts if he is to act freely. Free action need not *be* originated action, according to this view. Its freedom only requires originative *power*, since an action caused by unoriginated desires would

still be free as long as the agent was able to originate another desire instead, but refrained from doing so. In originating an act, an agent would be its source; the causal sequence leading to the action would begin in or with him.

Two sorts of arguments might be offered in support of the claim that someone must be able to originate his acts if he is to act freely. According to the first sort of argument, if someone cannot originate his acts, then he must intend to do what suits his unoriginated desires. But someone's unoriginated desires are either distinct entities in their own right or belong to a distinct "appetitive soul." So, either his distinct desires force him to intend to do what suits them, or the distinct appetitive soul to which they belong forces him to do so. Thus, he is forced to act as he does. Ultimately, though, the hereditary and environmental factors which produce the distinct desires or the appetitive soul to which they belong force him to behave as he does. Thus, he doesn't act freely, and so he doesn't meet the requirements for moral responsibility.

The notion that our desires are entities in their own right is obviously false. Desires cannot be conceived of except as something's states or psychic behavior. This sort of argument can only be sustained on the assumption that our unoriginated desires are the states or behavior of a distinct appetitive soul. However, the claim that our unoriginated desires belong to a distinct appetitive soul is contradicted by our experience of being the subjects of our desires in the same sense in which we are the subjects of our thoughts. Moreover, it seems to me that it is precisely someone's tendency to desire certain things with a certain intensity that makes him unique and interesting. When we think about someone, we think about his tendency to desire these things. We do not think that in directing our attention toward his desires we avert our attention from him. When

we love someone, loving *him*, we direct our attention toward what we take to be his good desires, and when we dislike someone we direct our attention toward what we take to be his bad desires. It is tempting to think that someone's good desires are truly his own states or behavior, while his bad desires aren't. However, to think this would be to obliterate the distinction between moral badness and psychological disorder, and would require us to conclude that guilt-feelings and repentance are manifestations of delusion. I conclude, then, that someone's bad desires are just as truly his own states or behavior as his good desires are.

Since someone's unoriginated desires are neither distinct entities in their own right nor belong to a distinct appetitive soul, the first sort of argument supporting the claim that someone must be able to originate his acts if he is to act freely must be rejected.

According to the second sort of argument that may be offered in support of this claim, while someone's desires are indeed his own, he is forced by his heredity and environment to have them. The hereditary and environmental factors that cause him to be such that he tends to desire certain things to a certain extent in certain circumstances thereby force him to desire these things to that extent in those circumstances. He is thereby forced to have the inclinations that he has, and to intend to satisfy these inclinations, and to act upon this intent if possible, unless he can originate an alternate act. Now, we can say that in tending to desire certain things with certain intensities in certain circumstances, someone is a certain *sort* of person. The question, then, is whether the things that make someone the sort of person he is thereby force him to act as he does. It seems to me that this cannot be so. If the things that make him the sort of person he is

thereby force him to act in the way in which he must act because he is that sort of person, he must already be a person capable of acting otherwise *before* he is made into that sort of person. How, though, can he be capable of acting otherwise before he is any *sort* of person at all? Even supposing there can be a person that is not a certain sort of person, a person that is not a certain sort of person would not have any preferences that could motivate him (or give him a reason) to act in one way rather than another, or, indeed, to act at all rather than to remain inactive. Such a person would be incapable of acting. But a person incapable of acting cannot be forced to act. It follows that whatever it is that makes a person the sort of person he is cannot be understood as forcing him to act as he does. But whatever it is that makes a person the sort of person he is would thereby be a prior cause of his act – a cause prior to any act or event of which he is the subject, or to his causing of anything. Someone's acting freely therefore does not require his being the originator of his act, or his being able to originate another.

Note that this argument does not depend upon the thesis that someone's heredity and experiences produce *him*, as the person he is. Perhaps God makes unformed persons and casts them down into the mold of heredity and experiences. If heredity and experiences do produce persons, as the sorts of persons they are, then since there is no person prior to this production they surely don't force a person into a given form. The things that produce a person *as* this sort of person don't thereby force him to be this sort of person. But even if persons precede their formation by their heredity and experiences, they can't be forced by their heredity and experiences to act as they do, for before they are thus formed they are incapable of acting.

Those who propose the origination of acts as a requirement for free action cannot coherently draw upon the ordinary notion of free action as unforced action. They must deny that someone acts freely when his act expresses any propensities whose causal antecedents all precede his own existence. Whether or not someone's personality is the product of his heredity and experiences is irrelevant to the question of whether or not he is forced to do what he does. So, perhaps those who propose that free action requires origination draw upon another ordinary notion of freedom as disentanglement. A free act would be one that is not tangled up in the causal web of nature and history, just as we might say that a butterfly released from a net has been set free. It seems appropriate to call this sort of freedom "libertarian freedom", so as to invoke the thought of butterflies and former prisoners "at liberty." (Note that it is the *act*, and not its agent, that is compared to the butterfly and former prisoner. The "causal web of nature and history", from which the act is detached, is compared to the butterfly net and jail from which the butterfly and prisoner are released.)

Insofar as our personalities are products of heredity and experience we do not act in libertarian freedom. Our libertarian freedom need not be complete, since our personalities might be partly, but not entirely, products of heredity and experience. Conveniently, those who urge that we have libertarian freedom as a result of having the ability to originate our acts are commonly known as "Libertarians".

It should be noted, however, that while libertarian freedom, as I have defined it, seems to be compatible with a purely random formation of intentions, the origination of acts is not compatible with such randomness. Origination entails libertarian freedom, but there can be libertarian freedom without origination. What seems to matter to most so-

called Libertarians, though, is not libertarian freedom itself, but rather the ability to originate acts. They care about libertarian freedom only in so far as it is the result of an ability to originate acts. They often show that originaive ability is what they really care about by insisting that *moral responsibility* requires what they call freedom. Only libertarian freedom resulting from an ability to originate acts could plausibly make a difference where moral responsibility is concerned. Random action (behavior caused by randomly arising intentions) is action for which the agent *cannot* be morally responsible. If we think that someone's intention has arisen randomly, we cannot regard it as expressing his personality, and so shouldn't like or dislike him as a person in considering it.

Here I have in mind an intention that arises randomly even with respect to its general character as, say, violent or helpful. Someone would certainly deserve credit or blame for intentions which in their main contours express his personality but whose details are random (for example, the intention to punch someone in the jaw, as opposed to punching him in the nose). Furthermore, to say that we should not like or dislike someone as a person in considering what we believe to be his randomly arising intention is not to say that we should not protect ourselves against people whose random activity is frequently dangerous, or cannot reliably be expected not to be dangerous.

The claim might be made that while the origination of acts is not required for freedom, it is still required for moral responsibility, since if someone doesn't originate his acts it isn't really *he* who acts. This claim might be based upon the notion that not only someone's unoriginated desires, but even the intentions caused by these desires, belong to an appetitive soul distinct from him. It isn't he who intends, then, but rather the

appetitive soul which intends within him, driving him along – or worse, an appetitive soul within which the true, inner person is encased, carrying him along. So, any liking or disliking of someone as a person in considering acts which are really his appetitive soul's acts would be misdirected. It could never be right for one to like or dislike *him* as a person in considering these acts, and so he couldn't be morally responsible for them.

What I have just said in opposition to the notion that an appetitive soul distinct from us is the real subject of our unoriginated desires applies here as well.

If moral responsibility for an act requires its origination or the ability to originate another, then our punitive practices should be completely transformed. We are far from sure that someone is in any given case the originator of his act. The continuation of our present punitive practices therefore involves great moral risk. For, it seems that unless we have a special guardian-relationship to someone that makes it our duty to contribute to the shaping of his behavioral dispositions, such as that of a parent to a child, it is wrong that we punish him for an act when we have serious doubts regarding his moral responsibility for it. This is not to say that it is wrong for us to defend ourselves against him or to attempt to deter him from repeating his harmful acts by punishing him.

However, our present punitive acts go far beyond these limited measures. We assume, at present, that someone should be punished for a past harmful act even when he is unlikely to repeat it. No doubt, this punishment has some deterrent effect upon *other* people who would otherwise be likely to act in that way in the future. However, it seems to me – and although this is only a “intuition” it is one that I am sure I share with many other people – that it would be wrong to punish such a person only for the sake of this deterrent effect upon other people, if it is very doubtful that he is morally responsible for the act in

question. If people must originate their acts or be able to originate alternate acts if they are to be morally responsible for them, then since we can't know whether or not they indeed originate their acts we never have the right to punish anyone.

Surely, this places a heavy burden of proof upon those who hold that someone is morally responsible for an act that he doesn't originate only when he can originate an alternate act. The more difficult a social and psychological transformation would be, the higher the standard that must be met for the justification of this transformation. Our acceptance of the proposal in question would require either the severe restriction of our punitive practices or the abandonment of a liberal political system and its replacement by one in which government has the role of parental guardian. If restriction is chosen, punishment would be administered only in order to deter the malefactor himself from the likely repetition of his misdeed, or in order to prevent him from committing one in the first place, when he is evidently intending to do so. Only the parent-child model of government's relationship to a conglomerate population would justify the punishment of one person in order to deter another, when the first person is very likely not morally responsible for what he has done. Either approach (the sharp restriction of our penal practices or the transformation of our political model) would be extremely difficult. It seems, then, that the standard of justification that the proponents of the proposal in question must meet is unreachably high.

I will now review four accounts of what free action and/or moral responsibility requires that do not include among these requirements an ability to originate acts. Three of these accounts are offered as accounts of free action, or of free action and moral

responsibility simultaneously. I will assess these first as accounts of free action, and then as accounts of moral responsibility.

Harry Frankfurt proposes that someone “acts of his own free will” when his desire to behave in a certain way causes him to behave in this way *and* he wants this desire to cause him to behave in that way.²⁵ Someone’s “will”, according to Frankfurt, is a desire of his “that moves (or will or would move) a person all the way to action.”²⁶ The qualification “or will or would move” suggests that someone’s will, when he knows that it is impossible for him to escape from his cell, might still be that he leave the cell. Thus Frankfurt’s “will” cannot be synonymous with “intention”, even if intentions are reducible to certain interrelations of desire and belief, as Frankfurt might hold. Someone cannot intend to do what he does not think that he can do. Frankfurt seems to use “act” to mean what I would call “mere behavior.”

I offer my stipulated distinction between the uses of “action” and “(mere) behavior” as a convenient way of indicating whether or not behavior is intentional; only intentional behavior is to be called “action”. Even if Frankfurt is right in holding that desires cause behavior directly, rather than by causing intentions which in turn cause behavior, as I hold to be the case, he might still distinguish between intentional and unintentional behavior by calling only the former “action”, as I do. It seems likely, then, that he would call someone’s behavior “intentional” only if it is caused by the person’s desire to behave in that way. So, Frankfurt might agree to call someone’s behavior an “act” only when it is caused by his desire to behave in that way. (Such a desire would be the person’s will.) Frankfurt might then say that someone acts *freely* when and only when

²⁵ “Freedom of the Will and the Concept of a Person”, 89.

²⁶ *Ibid.*, 80.

his behavior is caused by his desire to behave in this way and he wants this desire to cause him to behave in the desired way.

Suppose, then, that Frankfurt is right in holding that desires cause behavior directly, rather than doing so by producing the distinct motivating pro-attitude that I call “intention”, which in turn causes behavior. It seems to me that even if he is right about this, still, whether or not someone wants his desire to behave in a certain way to cause him to behave in this way has no relevance to the question of whether or not he is forced to act – that is, intentionally behave – as he does. It would have no relevance to this question because someone can fail to want his desire to behave in a certain way to cause him to behave in this way without being forced to want to behave in this way. He can even want his desire to behave in a certain way to *not* cause him to behave in this way without being forced to want to behave in this way. He is forced to behave as he does only if his behavior is contrary to his personality, which is his system of propensities to want certain things under certain circumstances. His propensity to want A is as much a feature of his personality as is his propensity to want his desire for A not to be effective. A supporter of Frankfurt’s proposal might suggest that someone’s propensity to want his desire for A not to be effective is somehow “deeper” than his propensity to want A. However, I don’t see what this additional depth could consist in, other than the fact that the desire that the desire for A not to be effective refers to a desire (it is a “second-order” desire). I don’t see how this “second-order” status of the desire that the desire for A not be effective can suffice to make the striving for A contrary to a personality which contains a propensity to want A along with a propensity to want the wanting of A to be ineffective. Nothing changes if the second-order desire is simply a desire not to want A

at all. I don't see why the propensity to have this desire should be privileged in the characterization of the personality over the propensity to want A. Both desires are equally characteristic of this personality. So, even if I were to accept Frankfurt's assumption that desires cause behavior directly, I would reject Frankfurt's account (in its adapted formulation) as an account of what someone's acting freely requires.

We can adapt Frankfurt's account of free action so that it fits my assumption that desires cause behavior only by causing intentions that in turn cause behavior. It then becomes the proposal that someone acts freely when and only when his behavior is caused by an *intention* that he wants to have. (It couldn't become the proposal that someone acts freely when and only when his behavior is caused by an intention that he *intends* to maintain in himself in some way, since such an intent is so rare that if this were what free action required we would rarely, if ever, act freely.) Again, though, and for a similar reason, it seems that someone's behavior can be caused by an intention that he doesn't want to have, or wants not to have, without his being forced to behave in this way. His desire not to have the intention that he has would only be one desire in the mix of desires that together produce his intention. This desire seems no more (and no less) expressive of his personality than his other desires are, including the desire or desires that he intends to satisfy. Since his intention expresses his personality, he isn't forced to act as he does. So, his not wanting to have, or wanting not to have, the intention doesn't make his act unfree.

It is true that the behavior that we call "compulsive" is behavior caused by an intention of a recurring type that the compulsive person probably doesn't want to have, and "compulsive" is synonymous with "forced". However, our calling this behavior

“compulsive” simply reflects our imaginative license: we imagine the person’s intention (or the configuration of desires that causes this intention) as something alien to him, because we think that this is how it must seem to him, given his desire not to have intentions of this type. A misleading picture of unwanted intentions as alien entities leads us to call his behavior compulsive, and so the fact that we call it this shouldn’t lead us to think that his behavior is unfree. To say that his behavior is free, though, is not to say that he is morally responsible for it. He probably isn’t.

Perhaps, though, Frankfurt’s account, adapted to fit my intention-based scheme, works as an account of moral responsibility. Perhaps someone’s being morally responsible for his behavior requires his wanting to intend to behave in that way. This proposal initially seems plausible, because someone might attempt to excuse himself from reproach or punishment by claiming that he hadn’t wanted to behave in the way in question. I reject it, though, because it is incompatible with the assumptions that guide our actual punitive practices. We seem to assume that unless we have a special relationship to someone, such as that of a parent to a child, that makes it our duty to positively contribute to the shaping of his behavioral dispositions, it is bad that we punish him when he is not morally responsible for his bad behavior. The rules governing the deliberations of juries regarding someone’s liability for punishment presumably codify our ordinary thinking about what makes it okay (not bad) that someone be punished, and our ordinary thinking about this should guide our inquiry. If juries don’t worry about whether or not someone wanted to intend to harm a person then, we should presume that his wanting to intend to harm the person isn’t a requirement for his being morally responsible for harming that person. And, in fact, juries *don’t* ordinarily worry about

whether or not someone wanted to intend to harm someone else when they deliberate over whether or not to punish him. They worry about whether he intended the harm, but not about whether he wanted to intend the harm. So, someone's wanting to intend to do something wouldn't be a requirement for his being morally responsible for doing it.

Juries may indeed be inclined to punish people less severely when they think that these people don't want to intend to do the bad things that they intend to do. However, this inclination does not show that we don't regard such people as morally responsible for their acts. It only shows that we regard them as having a less *intense* moral responsibility for their acts. That is, we think that they deserve less blame – that we should have a less unfavorable attitude toward them. Presumably, the extent of the severity with which it is okay, or even right, that we punish people for their acts is somehow proportional to the intensity with which we should dislike them as persons in considering these acts.

Gary Watson offers another proposal about free action that can be reframed as a proposal about moral responsibility. Watson proposes that one acts freely when “what one most values [is] what one is finally moved to get.”²⁷ It is his “valuing” of a thing that moves him to get it by means of his act, if he acts freely: someone acts freely when and only “his actions flow from his evaluational system.”²⁸ Watson's “valuing” might be interpreted as a special sort of “pro”-attitude toward the thing that is valued, rather than the belief that the thing has any real goodness or rightness. However, in invoking Plato as the inspiration for his proposal Watson speaks explicitly of “the good,” and of “judging good,” and in so doing seems to endorse a conception of goodness as a property that resides in things independently of anyone's attitude toward these things. Perhaps he

²⁷ See “Free Agency” [1975], in Ekstrom 2001, 92-105; 95.

²⁸ *Ibid.*, 101.

wishes to leave open the question of whether evaluative utterances are to be interpreted as claims or as expressions of attitudes. I am interested in the proposal that someone acts freely only insofar as he is motivated by his desire for whatever is best. (He would then do what he thinks best, for the sake of its being best.) For present purposes, I will attribute this proposal to Watson.

Does Watson have in mind that someone who acts freely must be motivated by a desire for what it is best *overall*, or by a desire for what is best for *him*? The first interpretation doesn't necessarily require that a free agent be an altruist, since he might stupidly misperceive private goods as universal ones. The second interpretation doesn't necessarily require that a free agent be selfish, either, since his helpful act may be what is best for *him*.²⁹ I will leave this question open.

It should be noted that Watson's proposal loses whatever plausibility it has as soon as it is adapted so as to reflect the language of "right" and "wrong", as opposed to "good" and "bad". It then becomes the proposal that someone acts freely only insofar as he is motivated by his desire to do whatever is right, under the direction of his judgment that a certain act would be right. This would be far too restrictive. Most of the cases in which we think that someone deserves blame for acting as he did (cases in which we therefore assume that he acted freely) are not cases in which we think that he acted as he did in order to do whatever is right. There is some plausibility to the claim that people always do what they do for the sake of the goodness of the act or of its anticipated consequence. (For example, they think that the pleasure it brings them is something

²⁹ Plato himself has Diotima winning Socrates' assent to the proposal that people "want the good to be *theirs* forever." *Symposium*, trans. A. Nehamas and P. Woodruff, in *Plato: Complete Works*, ed. John M. Cooper (Hackett Publishing Company: Indianapolis/Cambridge 1997), 206a. Emphasis added.

good.) There is no plausibility to the claim that people always do whatever they do for the sake of the rightness that they attribute to doing it.

We can align Watson's proposal more closely with my own picture of motivation if we say that someone's predominant desire to do whatever is best causes him to behave as he thinks best by causing him to *intend* to behave in that way. Someone acts freely only insofar as the intention that causes his behavior is caused by his judgment that it is best to behave in that way, together with his predominant desire to do whatever is best. This doesn't entail that an agent would regard his unfree act as bad. He might not be thinking about its value at all.

Watson sees behavior that isn't caused by a desire for whatever is best as caused by a passion or urge. He sees someone whose passion or urge causes his behavior as forced by the passion or urge to behave in that way. Apparently, then Watson sees someone's passion or urge as an alien entity inhabiting or impinging upon him. Or else, he sees someone's impassioned or urge-driven alter-ego as pulling and pushing him along. I therefore reject Watson's proposal as an account of free action. I don't see that someone whose intention isn't caused by such a judgment, but is instead caused by his passions or urges, would therefore be forced to have the intention that he has. I don't see this, because I don't see someone's passions and urges as entities alien to him, or as properties of an entity alien to him. I see them rather as attitudes or even continuing events of which he is the subject.

How does Watson's account of what makes an act free do as a proposal regarding moral responsibility? This would be the proposal that someone is morally responsible for his act only when he is motivated by his desire to do what is best. Again, such an

account is incompatible with the assumptions that guide our ordinary punitive practices. For one thing, it entails that we would think it wrong to punish anyone for his criminal behavior whenever he behaved in that way just because he was so inclined rather than because he thought it best that he behave in that way. Anyone whose bad behavior isn't motivated by the mistaken opinion that this behavior is good would be excused from moral responsibility for it, and moral responsibility is an apparent prerequisite for liability for punishment. But even with regard to those people who intend to behave in a way that turns out to be bad when this intention is caused by a motivating judgment that this behavior is good, this account of moral responsibility fails. Juries don't think that it is the causing of someone's criminal behavior by his judgment that this behavior is good that makes him morally responsible for this behavior, and therefore liable for punishment. They don't even think that this contributes to his moral responsibility. They don't inquire about it.

John Martin Fischer proposes that someone is morally responsible for his act when there is some possible state of affairs in which the psychic "mechanism" that actually causes his behavior might function in a way that involves his "recognition" of a "sufficient reason to do otherwise" and might as a result cause him to do otherwise.³⁰ But the psychic mechanism that operates in the case of acts motivated by what we might call "uncontrollable passion" is simply the causing of action by such uncontrollable passion, and therefore *can't* function in a way that involves his recognition of a sufficient reason to do otherwise. Someone driven by his sudden rage to commit a violent crime therefore couldn't, according to Fischer's account, be morally responsible for the badness of his violence. Yet a very large percentage of our punitive responses are to acts of

³⁰ *The Metaphysics of Free Will: An Essay on Control* (Blackwell Publishers: Cambridge 1994), 164-168.

violence caused by sudden rages, just as a very large percentage of our honor-conferring responses are to acts of heroism and generosity caused by sudden bursts of altruistic feeling. These are behavior-causing psychic mechanisms that simply can't function in a way involving the recognition of a sufficient reason to do otherwise, where this recognition would cause the person to do otherwise. Fischer's account therefore requires too major a transformation in our rewarding and punitive practices to be acceptable.

Susan Wolf's account of the personal requirements for moral responsibility is more satisfactory, and I will adopt something like it.³¹ Wolf speaks only of "responsibility", but clearly has in mind what I have been calling "moral responsibility." She seems to regard someone's acting freely and his being morally responsible for his act as entailing one another, and so presents her account as an account of what is required for *both* free action and moral responsibility. I don't think that action must meet her proposed requirements to be free (unforced), but I do think that it must meet requirements like these if the agent is to be morally responsible for it.

Wolf proposes that "what responsibility really requires is the ability to act in accordance with and on the basis of the True and the Good."³² Here Wolf speaks of goodness rather than of rightness. However, since she also says that "the ability we are concerned with might be described as the ability to do the right thing for the right reasons," it seems permissible to interpret her as holding that someone meets the personal requirements for moral responsibility just in case he is able to form reasonable beliefs regarding what it would be right that he do, and to act accordingly. In acting accordingly, he would presumably be motivated primarily by a desire to do whatever is right, under

³¹ *Freedom Within Reason* (Oxford University: New York 1990), 68-73.

³² *Ibid.*, 71.

the guidance of his reasonable belief regarding what it would be right that he do.

Someone's moral responsibility for an act would then require that he can be motivated primarily by a desire to do whatever is right, and that he can form reasonable beliefs about the rightness and wrongness of his possible acts.

Even if we interpret Wolf's account as focusing upon someone's ability to recognize the comparative goodness or badness (rather than rightness or wrongness) of the acts that he can perform, this account is superior to Watson's, which also focuses upon beliefs regarding what is good or bad. Watson holds that someone's moral responsibility for his act requires that his desire for whatever is best have moved him to perform it. Wolf holds that someone need only be *able* to do something for the sake of its being best, and able to see what is best. Wolf's account thus enables us to regard people who are driven to commit violent crimes by sudden rages as deserving blame for their bad acts, while Watson's doesn't. Her account passes the test of compatibility with our punitive practices. It seems that all beings which we think it okay for our courts to punish are able to recognize the real value of their possible acts and so are able to do what is in fact best for the sake of its being best. However, none of the beings which we think it wrong for our courts to punish, such as animals, small children, and psychotics, are able to do this. It seems to me, though, that someone's ability to recognize the comparative *rightness or wrongness* of his possible acts, rather than their comparative goodness or badness, is what directly matters here. His ability to measure goodness and badness would matter only insofar as the rightness or wrongness of an act depends upon its comparative degree of intrinsic and/or instrumental goodness or badness.

People can be morally responsible for their acts to a greater or lesser degree. For example, we ordinarily think that great stress reduces a sane adult's moral responsibility for an act. Wolf's proposal suggests that it might do this by reducing his ability to recognize the real comparative rightness and wrongness of his various behavioral options. He would then be less able to behave in the way that is in fact right, even when his desire to do whatever is right predominates.

Someone's being morally responsible for his act doesn't seem to require that he have been aware of other possibilities when his intention began. However, it seems to require his awareness of other possibilities while his intention endures. This awareness of alternate possibilities is entailed by the requirements for moral responsibility that Wolf proposes – that someone is morally responsible for his act when and only when he is able to see what he should do and able to do it for the sake of its rightness. He can see what he should do only if he can see that it is more right than anything else that he can do.

It is unclear what an *ability* to be motivated primarily by a desire to do whatever is right would involve. It seems that if anyone's desire to do whatever is right is too weak to motivate him, then for all practical purposes he is unable to be motivated by this desire. In this case, though, whenever someone isn't primarily motivated by it he *can't* be primarily motivated by it. So, perhaps Wolf simply means to require that someone have some desire to do whatever is right. Perhaps this is indeed a requirement for moral responsibility. However, I am inclined to regard it as a requirement for personhood. I have proposed that someone is morally responsible for X just in case everything is true of him that must reasonably be believed about him both by anyone who in considering X should like him as a person and by anyone who in considering X should dislike him as a

person. If having a desire to do whatever is right is a requirement for personhood, then it would be redundant to list it as a requirement for moral responsibility. But Wolf's other proposed requirement, that someone be able to form reasonable beliefs about the rightness and wrongness of his proposed acts, is insufficient. Such a person might have intentions, or even intention-causing desires, induced in him through hypnosis and the like. I would suggest instead, as an additional requirement, that the intention for which he is morally responsible express his personality – his system of propensities to have certain desires, of certain strengths, under certain circumstances. Someone who acts upon an intention that expresses his personality would be acting freely. So, someone's moral responsibility for his act requires that he act freely.

3. *Moral goodness and badness*

These, then, are the requirements for moral responsibility that I propose – the things that both the people who should blame someone for his intention *and* the people who should give him credit for it must alike reasonably believe about him: (1) his intention must express his personality, and (2) he must be able to have reasonable beliefs about the degree of his intended act's *rightness or wrongness*, as compared to that of other possible acts. I have been urging that free action is action expressing the agent's personality. If this is correct, then the first requirement for moral responsibility for acts is just the requirement that the agent have acted freely. His ability to have reasonable beliefs about the degree of his act's rightness or wrongness would entail an ability to have reasonable beliefs about the degree of its intrinsic and/or instrumental goodness or badness, since *one* of the things that can make an act right or wrong is that it produces more good or bad than the alternatives do.

I formulate the first requirement as the expression of *personality* by intention, rather than the expression of desires (in their overall configuration) by intention, because I am queasy about speaking of someone as unconsciously desiring something. Rather than thinking of someone as having enduring desires that pop into consciousness from time to time, I would rather say that he has enduring propensities to have these desires under certain general conditions. I have argued that the prisoner who talks under torture, and the person who leaps from the window of a burning building, do so unfreely because these acts don't express their personalities. So, they certainly aren't morally responsible for these acts. It took some examining to see why these acts don't express their personalities, since their personalities do, after all, include propensities to want to escape

from extreme stress in any way possible. It is much more obvious that if someone were directly caused (by a neurosurgeon or hypnotist, for example) to desire things that he has no propensity to desire, without having chosen this temporary induction of desires, he would not be morally responsible for intentions dependent upon these desires. It sounds odd to say that he would not have “intended” freely, but he would clearly have been forced to intend as he does, in the same way that Sisyphus forces his rock uphill. In any case, if he were to carry out his intention, he would be acting unfreely. Only when someone’s intention expresses his personality does he act freely.

As for what rightness and wrongness are, I will leave this an open question. I assume only that rightness and wrongness are real properties (as opposed to “X is right/wrong” type-utterances being expressions of positive or negative attitude), for I assume that people can really be morally responsible for things. If rightness and wrongness aren’t real, then my definition of moral responsibility, relying as it does upon a statement about what “should” be the case, makes no sense.

“Right” can’t simply *mean* “productive of at least as much good as any other option,” as opposed to being a distinct property that an act has in *virtue of its* productiveness of at least as much good as any alternate act would produce. It is easy to think of cases in which the act that at least *seems* to produce the most good isn’t right. Fantastic thought-experiments make this obvious: for example, Dostoevsky’s case in which the torture of a child would bring about utopia, or the case in which the extraction of an innocent man’s vital organs might save the lives of several people.³³ A real-world

³³ Philipa Foot, “Killing and Letting Die,” in *Contemporary Moral Problems*, 9th ed., ed. James E. White (Thompson Wadsworth 2009) 166-171, 166, reprinted from Joy L. Garfield and Patricia Hennessey, *Abortion: Moral and Legal Perspectives* (Amherst: The University of Massachusetts Press, 1984), 177-185; also Louis P. Pojman, “Strengths and Weaknesses of Utilitarianism”, in *Exploring Ethics*, ed. Steven M.

example would be that of Gauguin, who can present the world with the fruits of his artistic talent and inspiration only by abandoning his family. More prosaically, someone may think that it would be best, although wrong, for her to lie on her resume -- best because best for her and her family while not harming her prospective employer. It doesn't matter that it may also suit her desires to lie on her resume; the point is that she thinks it best, although wrong. Such cases also make it doubtful, although not impossible, that an act's productiveness of at least as much good as any alternative necessarily makes it right.³⁴

Even if all right acts produce at least as much good as any alternative, and all acts producing at least as much good as any alternative are right, it is still more convenient to speak in terms of rightness and wrongness than in terms of productiveness of goodness and badness. Furthermore, when we try to determine our duty, or deliberate over whether to do it, we use the words "right" and "wrong" in our thinking, rather than expressions explicitly indicating an act's greater or lesser degree of goodness-production. In such cases, our thought that one of the possible acts would produce the most good may be what makes us think that it is right. However, we ultimately focus on its rightness.

This is not to deny that we may think about what it would best to do, without considering the question of what it would be right to do. If a sense of obligation does not infuse our musings, we are probably not considering the question of what it is right to do. In the next chapter, I will argue that when we deliberate over whether or not to do something other than what we are inclined to do we necessarily deliberate over whether

Cahn (OUP 2009), 104-111, 109, reprinted from Pojman, *How Should We Live? An Introduction to Ethics*, (Wadsworth 2005).

³⁴ Pojman formulates "act utilitarianism" as the doctrine that "An act is right if and only if it results in as much good as any available alternative." *Ibid.*, 104.

or not to do what we think would be more right than what we are inclined to do. We would consider the intrinsic or instrumental goodness of the act in question only insofar as we think that the act's intrinsic or instrumental goodness is what makes it right.

Someone whose intention expresses his personality and who can have reasonable beliefs about his act's comparative rightness or wrongness is morally responsible for his intention. If he is an evidently sane and intellectually competent human adult, we may assume that he meets these requirements. Some further reasonable belief about him must then make it right that we like or dislike him as a person in considering his intention.

There seem to be three kinds of cases in which someone who meets the requirements for personal responsibility would deserve credit for his intention. These are the three ways in which someone would be *morally good* in having the intention that he has. Each way of being morally good can be given its own name, in very rough conformity to ordinary usage.³⁵

First, someone would deserve credit for his intention if he intends to perform an act that he reasonably believes to be right, and intends to perform it at least partly for the sake of its rightness. We can call this sort of moral goodness "conscientiousness".

³⁵ The following classification of types of moral goodness is similar to what Utilitarians and Deontologists alike have proposed. On the Utilitarian side, G. E. Moore's two types of intrinsically good "virtue" correspond to what I call conscientiousness and practical virtue (*Principia Ethica* [Buffalo: Prometheus Books, 1988], 177-180). Mill must have at least the conscientious sort of moral goodness in mind when he remarks that "...the motive has nothing to do with the morality of the action, though much with the worth of the agent" (*Utilitarianism, On Liberty, Considerations on Representative Government*, ed. Geraint Williams [London: J.M. Dent 1993], 19). On the Deontological side, D. H. Prichard applies the names "virtue" and "moral goodness" to Moore's two types of intrinsically good virtue ("Does Moral Philosophy Rest on a Mistake?", from *Mind*, XXI (1912), 487-499, in Joseph Margolis, ed., *Contemporary Ethical Theory: A Book of Readings* (New York: Random House, 1966), 47-66, 58, 53, 59-60). In proposing three species of "morally good" action, of which action motivated by the desire to do whatever is right is one, W.D. Ross divides the other sort of moral goodness into two separate categories (*The Right and the Good*, ed. Philip Stratton-Lake [Oxford: Clarendon Press 1930, 2002], 160). Kant only recognizes the conscientious sort of moral goodness. His "good will" is that of the morally good agent intending always to conform his activity to the moral law. Acts thought to conform to *the* moral law would ordinarily be called "right" solely in virtue of this conformity.

Strictly speaking, though, if we take ordinary usage as our guide, conscientiousness doesn't require that someone's beliefs about what is right and wrong be reasonable; it requires only that someone do what he thinks is right for the sake of this supposed rightness. So, moral *badness* can take the form of conscientiousness as well – the bad conscientiousness of someone who intends to perform an act that he *unreasonably* believes to be right for the sake of this supposed rightness, when the act is really wrong. There is a good conscientiousness and a bad conscientiousness. By “conscientiousness” I will mean good conscientiousness, however, unless the context indicates otherwise. The conscientious person's desire to do whatever is right would play a prominent role in the production of his intentions – and this is true of bad as well as good conscientiousness.

Second, someone would deserve credit for his intention to perform an act which he reasonably believes to be right, and which he would not intend to perform if he believed it to be wrong, when he intends to perform it at least partly for the sake of the *goodness* that he reasonably attributes to it. This goodness might be either intrinsic or instrumental.³⁶ I can't think of a better name for this sort of moral goodness than “idealism.” However, just as in the case of conscientiousness there can be a bad conscientiousness, so here as well there can be a bad idealism – the idealism of someone whose beliefs about what is good and bad are unreasonable and false. Such an idealism would be a form of moral badness. But again, when in speaking of idealism I do not clearly indicate that I have bad idealism in mind I will be referring to good idealism. The idealistic person's desire to do whatever is *best* would play a prominent role in the

³⁶ If someone intentionally brings about some state of affairs by means of a core act, then the bringing about of this state of affairs by means of the core act is his “greater” act. If he reasonably believes this resulting state of affair's goodness to outweigh the core act's badness (if any), then he reasonably believes the greater act to be good. It seems to me that the greater act's goodness would be intrinsic, while the core act's goodness would be instrumental.

production of his intentions, whether his idealism is bad or good.

Third, someone would deserve credit for his intention to perform an act which he reasonably believes to be right, and which he would not intend to perform if he believed it to be wrong, when he performs it (a) for its own sake, when acts of this kind tend to be right or very good, or (b) for the sake of something that is very good (although not for the sake of this goodness). It seems appropriate to call this kind of moral goodness “virtue”. Rather than having goodness itself as its goal it aims at very good acts of certain kinds, such as the helping of others, or at bringing about certain kinds of very good states or situations, such as scientific understanding, or the enjoyment of beautiful things.

Someone might object to my distinction between “idealism” and “virtue” that *whenever* we do what we are inclined to do we do so for the sake of the goodness of what we are doing, even if we don’t have the word “good” in mind. This claim that we always act for the sake of goodness might be motivated by the thought that the goodness of an act or of anything that we might hope to bring about through the act consists in its fulfillment of some necessary desire of ours, so that the *apparent* goodness of it would consist in its seeming to offer us what has ordinarily been described as “self-fulfillment.” That is, all that is required for something to seem good for us is that it seem to fulfill, or complete, us. According to this position, to think to ourselves that it would be “good” to do something is only to explicitly state what we already feel when we merely want to do it. I regard this position as quite plausible. If I were to accept it, I would have to abandon the distinction between virtue and idealism. There would only be two kinds of moral goodness: conscientiousness and virtue. “Idealists” would just be virtuous people who explicitly formulate the abstract idea of goodness (as, say, conduciveness to personal

fulfillment). Or rather, virtuous people would be thinking “idealistically” when they explicitly formulate for themselves what it is about an act that inclines them to perform it. Usually, virtuous people wouldn’t have any such formulation explicitly in mind. Nevertheless, its content would have to be implicitly present to them; they would have to somehow “feel” (or vaguely imagine) that the act in question promises fulfillment, or whatever it is that goodness really consists in.

Having noted the possibility of adjusting my proposal regarding the forms of moral goodness in this way, I will return to distinguishing between the idealism of someone who does what he does for the sake of its goodness and the virtue of someone who only does what is in fact best for its own sake. This distinction seems to reflect our ordinary thinking about kinds of motivation – that we only *sometimes* do things for the sake of their goodness, and at other times pay no attention to the value of what we are inclined to do.

Idealism and virtue entail an overriding conscientiousness in cases where what is reasonably believed to be the right thing to do cannot be done with a typically idealistic or virtuous intention – cannot be done either for the sake of goodness (the idealistic intention) or for the sake of something good (the virtuous intention). Someone’s idealism in having the intention that he has would entail that he is generally conscientious with respect to any possible act that seems to him to be *the* right thing to do, but which does not seem to him to be best. Similarly, someone’s virtue in having the intention that he has would entail that he is generally conscientious with respect to any possible act that seems to him to be the right thing to do, and which he wouldn’t perform for its own sake or for the sake of anything that might be brought about through its performance. So, for

example, when the idealistic person thinks that it would be *better*, overall, when the goodness of the consequences are considered, to spend his time helping Mike than to keep his promise to Tom, but he thinks that his keeping his promise to Tom is nevertheless the *right* thing to do, he will keep his promise to Tom for the sake of the rightness of doing so. Similarly, if we replace the idealistic person in this situation with a virtuous person whose specific virtue is that he loves helping people, so that he would otherwise intend to help Mike (which, let us assume, would be a very good thing for him to do), still, when he thinks that his keeping his promise to Tom is the right thing to do, he will keep his promise to Tom for the sake of the rightness of doing so.

Virtue is a more restricted sort of moral goodness than idealism, because someone can be virtuously inclined to perform a certain kind of good act for its own sake without being similarly inclined to perform other kinds of good acts for their own sakes. Helping others and thinking about the meaning of life are both good, but someone who is inclined to be helpful for its own sake need not be reflective. There are different kinds of virtue, each kind being distinguished by the kind of good thing that it aims at. There is only one kind of idealism. Idealism always aims at goodness itself, and goodness is the same everywhere. Someone *might* have all of the different kinds of virtue, but this is unlikely.

We can reasonably conclude that someone is conscientious when we see that his act is the sort of act that people usually perform only for the sake of its rightness (as would be the case with acts of uncomfortable promise-keeping and truth-telling). We may then conclude that he has a fairly strong desire to do whatever is right, and that this desire will reappear to motivate him in the future with unusual frequency, by figuring prominently in the configurations of desire that generate his intentions. Our reasonable

conclusion that he has a propensity to intend to do whatever he believes to be right, and that his beliefs regarding what is right are reasonable, makes our liking him as a person the right attitude for us to have toward him. We should give him credit for his intention.

We can reasonably conclude that someone is idealistic when it is evident that he consistently does things that are not only right, but also very good, and very good in a *variety* of ways. This is especially true when there are other, less good but more immediately or intensely pleasant things that it would have been no less right to do. We should then like him as a person in considering any intention of his that seems to fit into this general pattern of activity. At the same time, we would be considering the general pattern of activity into which this intention seems to us to fit. So, we should give him credit both for the individual intention and for his general pattern of activity. Note that there is no inconsistency here with my position that whenever we deliberate over whether or not to do something we aren't inclined to do, it would always be for the sake of its *rightness* that we would decide to do it. If someone intends to do only what he takes to be the best of the things that he thinks right *for the sake of their goodness*, this would be because he has a predominant desire to do whatever is best, given that it is right. Because of this desire's comparatively great strength, he is inclined to do whatever is not only right, but best, and as a result intends to do so.

It seems, though, that our consideration of a single act would never permit us to reasonably conclude that the agent is idealistic, since this act would only be one kind of good act, or would only be the bringing about of one kind of good thing. Our consideration of a single right act that is better, although less immediately or intensely pleasant, than other possible right acts would only allow us to reasonably conclude that

the agent *either* has the corresponding kind of virtue or is idealistic, and that he *probably* only has the corresponding kind of virtue, since this is a much more common occurrence.

Moreover, we would be able to reasonably conclude that he is virtuous or (creditably) idealistic only if we reasonably concluded that *he* reasonably believes the act right, *and* that he would not have performed it had he believed it wrong. His belief in the act's rightness would usually be implicit. He would declare it right if he were asked for his opinion, and would be able to defend this opinion if challenged to do so.

In the absence of evidence to the contrary, it is reasonable to assume that a sane adult's beliefs about what is right and wrong are reasonable. So, if we see that someone's act is right (that is, not less right than any alternative), we can reasonably assume that he reasonably believes the act right. Evidence that this is not so might include his performance of an act that seems wrong, when past observation and/or trustworthy testimony suggest that he is conscientious. In this case, we would have to consider the possibility that he unreasonably believes the act to be right.

Reliable evidence regarding someone's past behavior is also required if we are to reasonably conclude that someone would not have performed an act had he believed it wrong. Suppose someone gives a beggar money that he has promised to return to a lender by a certain hour. Giving money to beggars is in general good, but the rightness of this particular charitable act is doubtful, given the donor's promise to return the money. We should conclude that he would not have performed the act if he thought it wrong only if we have seen him keeping past promises when it was difficult for him to do so, or have heard trustworthy reports that he has done so. We cannot reasonably conclude, on the basis of our observation of a single act that the agent would not have performed it if he

had thought it more right that he do something else.

It is reasonable to assume, in the absence of evidence to the contrary, that someone has a normal degree of concern with doing only what is right – that he would at least avoid whatever he believes to be *very* wrong. And, as I have already noted, we can reasonably assume, in the absence of evidence to the contrary, that his beliefs about what is right and wrong are reasonable. So, we can reasonably conclude that someone whose act is not only right but better than another equally right and more easily-performed act is to *some extent* virtuous, and perhaps idealistic.

Someone is morally good in intending to do something when he is conscientious or idealistic or virtuous in so intending. Someone who meets the requirements for moral responsibility is morally *bad* in intending to do something whenever he intends to do something which can only reasonably be believed to be wrong. He is morally bad in intending to do something even when he *unreasonably* believes that what he intends to do is right, and when he intends to do it for the *sake* of this rightness that he *unreasonably* ascribes to it. He may therefore be conscientious (in intending to do something for the sake of its supposed rightness) as well as morally bad. This is the bad conscientiousness that I mentioned above. He may also be idealistic in a bad way, as I have noted – doing what is wrong for the sake of the goodness that he *unreasonably* and *falsely* ascribes to it. If he does what is wrong for the sake of the goodness that he *reasonably* ascribes to the act, then he is idealistic while morally bad, but his idealism itself would not be bad. It seems to me that if he does what is wrong for the sake of goodness that he *unreasonably* but *correctly* ascribes to the act, this would be a bad idealism, since his ascription of goodness to the act is only accidentally correct. A morally bad person may also be the

opposite of the virtuous person – he may be vicious, performing acts for their own sake or for the sake of what he thinks he will thereby bring about, when these acts or consequences are intrinsically bad as well as wrong. However, it doesn't seem that someone can be "wickedly" idealistic, having badness as his goal in the same way that the idealistic person (whether his idealism is good or bad) can have goodness as his goal. Nor does it seem that someone can be wickedly conscientious, having as his goal wrongness itself, or the doing of whatever is wrong.

Someone who performs an act only for the sake of what he correctly but *unreasonably* believes to be its rightness does so with a morally indifferent intention, it seems to me. For example, someone may rescue a drowning girl only because she is left-handed, and he believes that someone's left-handedness is all that makes rescuing him right. It doesn't seem that anyone should like him as a person in considering this act. It may also seem that when someone performs an act only for the sake of what he correctly but unreasonably believes to be its rightness, his intention may be morally *bad* rather than merely morally indifferent. This may seem to be so when he believes that what makes the act right is that it is likely to have some consequence that he unreasonably believes would be very good, when in fact all that can reasonably be believed about this consequence is that it would be very bad, and/or that any act performed primarily for the sake of this consequences would be wrong. There is an ambiguity here, though. His morally bad intention is an intention to do something more extensive than the act that is right. Act A (say, the rescue of a drowning girl) is right, but he intends to bring about consequence C (say, the annihilation of all redheads) by means of A. The bringing-about of consequence C by means of A is a greater act, B. He doesn't intend to perform A for

the sake of its rightness. Rather, he intends to perform B for the sake of what he *incorrectly* as well as unreasonably believes to be B's rightness, when it cannot reasonably be denied that B is wrong.

Someone who meets the requirements for moral responsibility, and who would not intend to perform an act that he believes to be *very* wrong, would be neither morally good nor morally bad, but merely morally decent, in intending to perform a right act for its own sake, when the act is neither good nor bad, or in intending to perform it for the sake of something good that he expects to bring about through it. We may assume, in the absence of evidence to the contrary, that someone would not intend to perform an act that he believes to be very wrong. The case of Huckleberry Finn, who helps Jim to escape from slavery while believing this is very wrong, also suggests that someone who believes an act to be very wrong when in fact it can only reasonably be believed to be right (so that his belief that it is wrong must be unreasonable), and who intends to perform it for its own sake when it is good, or for the sake of some expected good consequence, is at least morally decent, if not morally good, in so intending. (Here, again, someone who intends to perform A for the sake of consequence C may also be understood as intending to perform B, the bringing about of C by means of A.)

Someone's moral decency in intending something still entails his moral responsibility for this intention, since the moral decency of his intention requires that it express his personality and that he be able to form reasonable beliefs about the rightness and wrongness of possible acts. Our ordinary assumption is that someone may be morally responsible for his intention or act even when he deserves no credit or blame for it. This assumption is shown by the fact that even when we aren't sure whether

someone's decision is morally good, morally bad, or merely morally decent, we say that he will be morally responsible at least for the intended consequences of this decision.

I have been speaking in an odd way in saying that someone is morally good, bad, or decent *in* intending to do something. I have been speaking in this way in order to emphasize that someone's intention is not an entity distinct from him. Someone's intention is his *intending*, a state of his own being, or perhaps an event of which he is the subject. His having an intention is his being in the state of intending, or his being the subject of this event of intending to do something. So, the moral goodness or badness of someone's intention is only *his* moral goodness or badness in so intending. The moral goodness or badness of someone's act is only his moral goodness or badness in intending it.

To say that someone is morally good, bad, or decent in himself, apart from his intending anything in particular, is to say that he tends to have the kinds of intentions that I have been discussing. We cannot see someone as idealistically intending something without also seeing him as "overridingly" conscientious in himself. We see him as tending to intend to do what he reasonably thinks best, for the sake of this superlative goodness, *except* when he reasonably believes something else to be the right thing to do, in which case he will do the latter for the sake of its rightness. Similarly, we cannot see someone as virtuously intending something without seeing him as overridingly conscientious in himself. We see him as tending to intend to do good things for their own sake except when he reasonably believes it would be more right to do something else.

What we must reasonably believe about someone, then, if we should like him as a person in considering his intention, is that he is morally good in intending to do what he

intends to do, whether he conscientiously, or idealistically, or virtuously intends it. What we must reasonably believe about someone if we should dislike him as a person in considering his intention is that he is morally bad. Our belief that he is morally good or morally bad encompasses and extends beyond the beliefs that we should reasonably have about him if we attribute moral responsibility for his intention to him – that his intention expresses his personality and that he is able to have reasonable beliefs about the rightness or wrongness of possible acts. His moral goodness in so intending entails that he meets the requirements for moral responsibility, but is not entailed by his having met these requirements.

Someone deserves credit for his morally good intention or act, and blame for his morally bad intention or act. However, things seem to be a bit more complicated where the crediting or blaming of someone for the *consequences* of his act is concerned. Here, the usefulness of the distinction between core and greater act that I proposed in the first chapter becomes apparent. It seems that someone deserves credit for the consequences of his core act only if they belong to his morally good greater act – only if he intended to bring them about by means of his core act. So, someone can deserve credit only for his intention or (greater) act. He can't deserve credit for any of the consequences of his greater act. This may sound odd, but only because an unduly restrictive notion of someone's (greater) act is entertained. Suppose that a well-educated young woman, motivated by a reasonable sense of duty, intends to improve the prospects of the working-class, neglected children next door by tutoring them and taking them on outings to museums and concerts, and that as a result of her intention they end up getting professional jobs. We can understand their successes as belonging to her greater act, and

see her as deserving credit for them, since these successes are encompassed within the general idea of the improved prospects that she intended to bring about.

On the other hand, someone whose act is morally bad may deserve blame for consequences of his core act that do not belong to his greater act, but only if he could not reasonably have been unaware that his core act would very likely result in them, and if he could not reasonably have regarded the intentional bringing-about of these consequences by means of his core act as anything other than wrong. So, if our young woman unreasonably disregarded the probability that her meddling would result in the father's becoming romantically interested in her and in the mother's becoming therefore doubly resentful of her, resulting in angry relations between the parents as well as in retaliatory mistreatment of the children by the mother, then she would be to blame for, say, the resulting separation of the parents and the increased delinquency of the children. Consequences for which someone should be blamed would include things which, strictly speaking, are not caused by his act but which would not have occurred had he done something else. Thus, "sins of omission" are accounted for.

The credit or blame that someone deserves for the good or bad consequences of his intentions is identical with the credit or blame that he deserves for these intentions themselves. (Where he deserves credit for these consequences, they would belong to his greater act, but where he deserves blame for them they would not necessarily belong to it.) However, someone might still deserve rewards and punishments of a magnitude proportional to the overall goodness of the consequences of his morally good intentions and the overall badness of the consequences of his morally bad intentions; he should perhaps not be rewarded or punished at all for mere intentions (however morally good or

bad) that have no respective good and bad consequences. The economy of reward and punishment is distinct from, although related to, the economy of credit and blame, and belongs to the sphere of legal, as opposed to moral, responsibility.

When people correctly and reasonably believe that someone's act is morally good or bad, and like or dislike him as a person in considering his act, we can conveniently represent this by saying that they *respond* to his act by liking or disliking him as a person. Strictly speaking, they would be responding to the *act*, while focusing their attention upon his goodness or badness in intending it. However, we often speak of a response to someone's act, when we focus our attention upon him, as its agent, as though it is a response to *him*. This may be why we call *him* "responsible" for his act when we think that our response to the act should consist in our liking or disliking him as a person.

It seems that the credit or blame that we should give someone for his good or bad intention is proportional to the moral goodness or badness that we attribute to him. The greater the sacrifice of non-moral inclination entailed by someone's conscientious or idealistic intention to do what he reasonably believes to be right for the sake of its rightness or goodness, the more morally good he is. The better the thing someone virtuously intends to do or bring about for its own sake, when he reasonably believes the act in question to be right, the more morally good he is. (I assume that his intention expresses his personality, rather than being induced in him by hypnotic suggestion and so forth.) The more wrong someone's act, when he meets the requirements for moral responsibility, the greater his moral badness.

Crediting or blaming someone for his act consists in liking or disliking him as a person while (and as a result of) considering it. I have suggested that the liking or

disliking of someone as a person might involve a readiness or reluctance to become personally and/or communally involved with him. So, the more moral goodness or badness we reasonably attribute to someone, the more ready or reluctant we should be to become involved with him. That this is so is supported by the consideration that when we believe that someone has acted upon an intention that is very morally wrong (that is, has performed a very wrong morally wrong act), we are inclined to sever our communal ties with him altogether, by imprisoning or even executing him. A reason for our disinclination to do so when he has merely had a morally wrong intention, but hasn't acted upon his intention must be found. I doubt that our disinclination to punish someone for the mere intention that we attribute to him arises only from our uncertainty as to whether he has indeed so intended.

We may dislike or dislike someone as a person in different ways in considering his act. These differences seem to be functions of the different kinds of morally good or bad intention that we attribute to him. I have distinguished three basic kinds of morally good intention – that of the conscientious person, who intends to do things for the sake of their rightness, that of the idealistic person, who intends to do or bring about things for the sake of their goodness, and that of the virtuous person, who intends to do or bring about things that are in fact good. Our liking of the conscientious person takes the form of a special sort of respect. Our liking of the idealistic person takes the form of a special sort of admiration. Our liking of the virtuous person takes various forms, depending upon the character of the good that is his ultimate goal. We feel gratefully appreciative of an exceptionally generous person, for example. As for the ways in which we may dislike someone as a person in considering his intention when we believe it to be morally

bad, these seem to correspond to the kinds of wrongness that we attribute to his acts. So, we feel contemptuous of someone whom we take to have acted dishonestly, when we think that he meets the requirements for moral responsibility. We feel resentful of someone whom we take to have acted harmfully, when we think that he meets these requirements.

Our being morally responsible for intending to do something does not require that we originate this intention, or that we be able to originate an alternate intention. However, introspection suggests that we would have some special sort of moral responsibility for an intention that we originate, or for one that we have, but do not originate, when we can originate another one. I will address this question in chapter six. The proposals that I offer there will depend upon the specific account of the origination of intentions that I develop in chapters four and five.

4. *An event-causal account of origination*

The idea of someone's "origination" of an act is really just the image of the act as coming out of him, of him as its source, its starting point. One proposal regarding the reality that this image might represent is that someone who originates his act is the first cause of the act. This is the "substance causation" model of act-origination. Substance causation is the causing of events by substances. An agent is a substance, and his act is an event, so his causing of his act is substance causation.

Substance causation contrasts with the causing of events by other events, or "event causation". According to the dominant event causation model of how acts occur, the onset of a desire or belief, or some other psychic event, causes an act. It is generally thought that someone's origination of an act cannot occur through an event-causal process. The following argument leads to this conclusion. A psychic event that causes an act must either be caused or occur randomly. If it occurs randomly then it cannot constitute an agent's origination of anything. A subatomic particle would as justifiably be said to originate its "quantum leap." If it is caused by some prior event, and this prior event in turn by some prior event, then eventually this causal series must either begin with some random event of which the agent is the subject, or proceed from events of which the agent is not the subject. Either way, the agent cannot be imagined as originating his act. It seems to follow that an agent's origination of his act must be an instance of substance causation.

Someone who holds that substances can cause events might hold that substance causation can be either origination and non-origination.³⁷ Either way, a substance would

³⁷ This is Richard Taylor's view, although he speaks of "free" acts rather than of originated ones (*Action and Purpose*, 127-128).

cause an event. If the substance is an agent, and the event is the agent's act, the agent would cause his act. In cases of non-origivative substance causation, the substance would be caused by something else – probably by another substance – to cause the event. In cases of origivative substance causation, nothing else would cause the substance to cause the event.

The uneconomical character of a theory of causation according to which there is both event causation and substance causation is a reason to reject the proposal that there is substance causation. So, if someone's origination of an act must be substance causation, and non-origivative causation must be event causation, this would be a reason to reject the proposal that people can originate acts. But if there can be non-origivative substance causation, then perhaps all causation can be understand as substance causation. Instead of thinking that the impact of a hammer upon a nail causes the descent of the nail into wood, we can think that the hammer causes the nail's descent.³⁸ If so, then the proposal that people can originate acts would not be objectionable because uneconomical.

It seems to me, though, that what seems to be non-origivative substance causation is just event causation presented in a way that characterizes the causing event as the alteration of a substance. Either an event causes a substance to cause another event, or a substance causes another substance to cause an event. If event A causes a substance to cause event B, event A must alter the substance in such a way that the substance must

³⁸ According to Taylor, one feature of “metaphysical speculations involving cause developed prior to the rise of experimental science is that “an efficient cause was almost thought of as an object or substance... rather than as an event, process, or state” (Ibid., 11). Yet Taylor ends up holding that the only substance causation is an agent's causing of his act; all other causation is event causation. Timothy O'Connor, however, seems to endorse a universal substance causation picture, although he still calls non-origivative causation “event causation”: “It is because of the intrinsic properties of objects that a given particular has a given type of causal power.... Where the form of agent- and event-causal capacities differ is in their *exercise*. The exercise of event-causal capacities conforms to tendencies of some measure (the limiting case being deterministic). In the agent-causal case this is not necessarily so, as a conceptual matter” (*Persons and Causes*, 73).

cause event B. But then the alteration of the substance causes event B, and since this alteration is itself an event, we have event causation. If substance A causes substance B to cause an event, then again substance A must alter substance B in such a way that substance B must cause the event, and then, again, since this alteration is an event, we have event causation.

Perhaps we need not conclude that non-origivative causation is “really” event causation which is presented *as though* it were substance causation. Rather, we might hold there is a causation that may be described either as substance causation or as event causation. However, the same argument that leads to the conclusion that someone’s origination of an act cannot occur through an event-causal process also leads to the conclusion that someone’s origination of an act cannot occur through a causal process that may be described as event causation, even if it may also be described as substance causation. And, the difference between a causal process that can be described either as substance causation or as event causation and one that can only be described as substance causation seems sufficient to reintroduce a distasteful lack of economy into the theory of causation. So, if we hold that the origination of acts cannot be described as an event-causal process, we have a reason to reject the proposal that agents can originate acts.

We might as well call causation that can be described either as event causation or as substance causation “event causation”, and causation that can only be described as substance causation “substance causation”. So, if we hold that the origination of acts cannot be described as an event-causal process, but also hold that acts may be originated, we must hold that an agent’s origination of an act is the only substance causation there is. So, in saying that someone causes his act, we don’t need to add that he isn’t caused to

cause it. It may be assumed that he is its first cause. Ordinary speech certainly doesn't require that we indicate whether or not someone is caused to cause his act, when he causes it. In ordinary speech we never say that someone causes his act. We would say, then, there are two ways in which events may be caused: they may be caused by other events, or they may be caused by agents.

Even if this uneconomical rupturing of causation into two domains, causation by events and causation by agents, were acceptable, the assertion that someone causes his act without being caused to cause it is obscure. When it is said that no prior event, such as the onset of a desire or of a belief regarding how a desire might be satisfied, causes someone's act, but that *he* causes it, the confident emphasis on "he" may lead us to assume that there must be *some* significant content here. But what can this content be?

I don't see what the statement that someone causes his act can mean other than that, given his powers and given that external circumstances permit their activation, he must act in this way – that he can't exist in these circumstances with the powers that he has without acting in that way. The necessity that is invoked by saying that he "must" act in this way is not a logical necessity – not the necessity of an event whose non-occurrence is unthinkable. It must be a causal necessity, then. (When we speak of an event's "causal" necessity, do we mean anything more than that the necessity is not logical, that the event's non-occurrence is unthinkable? Non-logical, or causal, necessity seems unanalyzable, something primitively recognized.) So, his existence as something endowed with certain powers causes his act.

But he is his act's *first* cause only if nothing other than him causes his existence. To be his act's first cause, then, his existence must either be necessary or uncaused. But

if his existence is uncaused, then unless it is beginningless he must have randomly popped up out of nothingness. In this case, the image of him as his act's *originator* would be lost, for his act would ultimately have popped *with* him out of nothingness. Nothingness, rather than the agent, seems to be the act's source. So, if we take this approach to understanding someone's origination of an act, we end up requiring that he be a necessary being, something that *must* exist as the sort of thing it is. But this won't do, since we want to think of *ourselves* as able to originate our acts, and we are evidently not necessary beings.

We do not ordinarily say that someone causes his act. However, we do ordinarily say that someone causes an event. For example, we might say that someone made us laugh. Since someone's own act is an event, it might be thought that our ordinary talk of the causing of events by persons gives us license to conclude that someone can cause his own act. However, our ordinary talk of someone's causing of an event seems only to be another way of saying that a prior event of which he is the subject causes it. His *antics* made us laugh.

The foregoing considerations need not lead us to reject the proposal that we can originate our acts, though. We would only have to reject this proposal if it were true that the origination of act cannot be an event-causal process. But I think that the origination of an act *can* be an event-causal process. That is, I think that the image of origination, of someone as the source of his act, can be explicated in event-causal terms. This can be done if we understand the origination of act A as a beginningless chain of causally-linked originative acts that terminates in A, with the originating agent as the subject of each one

of the infinite number of originative acts in the series. No other substance would be the subject of an event prior to his participation in this causal series.

We should not accept infinite regressions unless we have to. Here, the need to accept an infinite regression arises only from a prior commitment to the idea that we can originate acts. This commitment does not require that we accept the infinite regression that would be entailed if we were to think of the causing of one event by another as being itself an event that the first event causes. We can think of the causing of one event by another as a relation between the two events that is no more an event in its own right than are the relations of “above” and “after” that can also hold between events. What seems to be the *occurrence* of the causal relation is only the occurrence of the caused event. The verbal form of the word “cause” doesn’t compel us to regard causation as an event, since there are other relations that aren’t events which are also indicated through the use of verbs – for example, “includes” and “precedes”.

If someone’s ability to originate an act is morally significant, then his origination of an act must itself be an act. The moral significance of this ability seems to generate our interest in it. So, I conceive of someone’s origination of an act as being an act in its own right – what might be called an originative act. Someone’s originative act causes the originated act to occur. It does so by causing the onset of an intention that in turn causes the originated act to occur. *Someone’s originative act is his act of forming an intention*, where (as we will see) this act is caused by his intention to form an intention, and this intention to form an intention is caused by his prior act of forming this intention to form an intention, and so on backwards in an infinite regression of intention-forming acts.

I have proposed a crude model of the relationships between the motivating pro-attitudes that I call “desires” and “intentions”, following what I take to be an ordinary use of these terms. A desire for something is a motivating pro-attitude toward it, where the desired thing is the *ultimate* object of this pro-attitude. Someone’s various desires together cause him to intend to do what he thinks will contribute to the satisfaction of his various desires to an extent roughly corresponding to their various strengths. I find it convenient to express this thought by saying that the configuration of his desires *inclines* him to do something. His inclination causes him to intend to do what he is inclined to do. He doesn’t necessarily have any *desire* to do what he is inclined to do, and what he therefore intends to do. If he did have such a desire, he would intend to do it “for its own sake,” as well as for the sake of attaining other things that he wants.

If intentions must be caused by desires, then someone’s origination of an intention would require his origination of a desire, or his origination of the strengthening of a desire. Someone might originate an intention to do something by originating the onset of a desire strong enough to alter the balance of his desires, so that the conglomeration of his desires causes him to intend to do what he believes will satisfy that desire to a certain extent. Or, he might originate his intention by originating the strengthening of an existing desire to the point where the same effect is produced. If he believes that the desire can be satisfied to a greater or lesser extent in various ways, his coming to intend to satisfy it in this rather than that way would depend on the strengths of his other desires, compared to the strength of the desire that he originates, or whose strengthening he originates. The stronger the originated desire, or the more it has been strengthened by his

originative act, the more he will intend to satisfy it by doing what he thinks will satisfy it to this greater extent.

My proposal, then, is that someone's originative act causes the onset or strengthening of a desire, and that this shifts the overall balance of his desires so that together these desires cause him to intend to do what he otherwise wouldn't have intended. If his originative act causes the onset of a new desire for X, then the rebalanced array of his desires causes him to intend to strive for X with a vigor roughly proportional to the strength of this new desire, as compared to that of his competing desires. If his originative act causes the strengthening of an extant desire for Y, then the rebalanced array of his desires causes him to intend to strive more vigorously for Y.

It may be objected that we don't deliberate over whether or not to want to do something; rather we deliberate over whether or not to do it. But, the objection goes, it is when we deliberate over whether or not to do something that we are most inclined to think that if we decide to do it we will have originated this decision. So, it isn't plausible that we would originate an act by originating a desire, or a desire's strengthening. My response is that it is only in our *moral* deliberations over whether or not to do the *right* thing that we really feel that we will have to originate the decision to do it if we are to do it. (I will be defending this proposal later.) Put in more ordinary terms, it is only then that we feel that we will have to *make* ourselves decide to do something, if we are to do it. We focus, then, on the possible *decision* to act, rather than on the possible act. But we feel that we will make ourselves decide to perform the act *by* making ourselves want, or want more intensely, one special thing that we think we can get only by performing the act – namely, that we do whatever is right. It seems to me that when doing whatever is

right is singled out as the only thing that we ever make ourselves desire, or make ourselves desire more intensely, in making ourselves decide to do something, it is no longer especially implausible that we would originate a decision by originating a desire or its strengthening. Any remaining implausibility is only the basic implausibility of anyone's originating an act.

I am proposing that someone's origination of his intention to behave in a certain way is itself an act -- something that he does intentionally. I must therefore hold that this origination act is caused by his intention to perform it (here I am thinking of intention as a continuing event that itself causes the intended behavior). His origination of his intention to behave in a certain way is caused by his intention to originate his intention to behave in this way. Since his origination of his intention to behave in a certain way is an *originative* act, his intention to originate his intention to behave in this way would itself have to be caused by his origination of this intention to originate an intention to behave in this way, and so on in an infinite causal sequence of events, each event having the same person as its subject.

But if someone's final intention (to move his limbs or direct his thoughts) is caused by the configuration of his desires, then it seems that for consistency's sake I ought to hold that someone's intention to originate an intention would also be caused by the configuration of his desires. This configuration of desires would have to feature a prevalent desire to originate this intention to originate an intention. If I were to hold this, I would have to hold that this prevalent desire to originate the intention to originate an intention, or its strengthening to the point of prevalence, is *itself* originated. Otherwise, the final intention couldn't be originated. So, we have to insert desires into the infinite

string of originative events: there is an origination of a desire to originate an intention to originate the final behavior, and this desire, in its prevalence, causes the corresponding intention. Beyond the consideration of consistency, we have this reason to introduce desires into the infinite originative sequence: someone's prevalent, natural, non-originated desires would exert a force *opposed* to his origination of a new configuration of desires in which these once-prevalent desires would no longer be prevalent. Suppose, for simplicity's sake, that someone's only natural desire is for pleasure. In this case, it seems that his desire for pleasure would oppose his origination of a stronger desire to do whatever is right. So, in order to originate this stronger desire to do whatever is right he would have to originate a prior desire to originate it, a prior desire that is stronger than his desire for pleasure. But then his desire for pleasure would also oppose his origination of this prior desire. Now a strong desire to originate *this* desire is needed! So, we must accept an infinite regress of desires to originate desires to ... originate an intention to perform the final act. This infinite regress causes the simultaneous infinite regress of intentions to originate intentions to... perform the final act.

These infinite regressions of desire and intention may be regarded by opponents of the claim that we can originate our acts as a demonstration of its absurdity. I am not quite ready to take the invulnerable, because conversation-ending, position of believing in the origination of acts "*because it is absurd.*" (Perhaps it is only silly, given the infinite regressions that it seems to entail, rather than logically absurd.) However, these infinite regressions are not quite so unacceptable as they might at first seem to be. First, since I do not hold that someone is the first cause of what he originates, as the substance-causation model of act-origination would have it, I do not face the problem of how to

place him at the beginning of an endless chain of causings of causings of his behavior or behavioral desire or intention. He is the subject of every event in an infinite chain of events causing events, including the final origination of the behavior or behavioral desire or intention in question. Second, we need not worry that we would have to be aware of every desire and intention that we have, yet are never aware of this infinite multitude of desires and intentions to originate intentions. All of these desires and intentions would be simultaneous, and a desire or intention to originate a desire or intention to originate a desire or intention would look pretty similar to a desire or intention to originate a desire or intention. I imagine the regression as being like a collapsible telescope made of transparent (and colored, or at least light-reflecting, and therefore visible) tubes: when the sections have been collapsed into each other and are viewed from the side, they all blend into each other. (The final desire or intention corresponds to the outermost section of the collapsed telescope.)

Nevertheless, we still face the difficulty of explaining why the whole infinite series should occur in the first place – a difficulty that we do not face when considering a series of events that does not include any personal originations of anything.³⁹ The occurrence of any series of events, no matter how long, that does not include personal originations of events can be explained by invoking the notion of natural law. Natural law cannot be invoked to explain someone's origination of an intention. It might be proposed that the notion of *purpose* can help us to achieve the understanding that we seek. The explanation of someone's origination of an intention is simply that he does it

³⁹ See Peter Van Inwagen, "Free Will Remains a Mystery", in James E. Tomberlin, ed., *Philosophical Perspectives*, 14, *Action and Freedom*, 2000 (Oxford: Blackwell Publishers Inc., 2000), 1-19, at 16-17.

in order to achieve such-and-such: for example, in order to do whatever is right.⁴⁰

However, this proposal is unsatisfactory, since we want to know why someone originates his intention in order to reach this goal, rather than beginning or continuing to intend to do what he is inclined to do. Only with his origination of the intention does this goal become *his* goal. He *makes* it his goal, taking on a purpose that he didn't have before. Why he does he do so? Here all I can only urge that a superhuman mind – the divine mind – might understand this occurrence, and so it need not be seen as being in principle random (that is, incomprehensible). This assumes that the notion of origination that I have developed is coherent. It seems to me to be coherent, although obscure. We can't see what it is that God would think, see, or grasp in understanding origination; if we can say this, we would already understand it ourselves.

It might be objected that even if God can understand why a person originates an intention, it cannot be understood by anyone why the origination should occur now rather than earlier or later.⁴¹ Here a plausible line of response would be that once the general features of the effect (someone's intention to do something) are sufficiently explained, the inexplicability of the more particular features of the act, including the moment at which it occurs, would not limit the agent's origination ability in any very regrettable way. Given that someone's origination of an intention is in principle understandable (by *someone*), and given that there is a certain time-range within which his intention to do something would *have* to begin, its beginning at one precise time rather than at another

⁴⁰ Richard Taylor proposes that the origination of an act is explained by the agent's "purpose" in performing it: he causes this act rather than another to occur because this act suits his purpose. See his chapter "Explanation and Purpose", in *Action and Purpose*, 203-223.

⁴¹ Carl Ginet's objection to the substance-causation model applies just as well to my own event-causation model: "A merely enduring thing as cause lacks the features needed to make it capable of explaining the particulars of the [behavior-causing] mental occurrence. It cannot, for example, explain its timing" (*On Action*, 13).

might as well be random. (We aren't surprised to find ourselves originating an intention at one moment rather than another because we weren't expecting to originate it at all. An expected intention can't be originated.)

If causation is an event, and all events are caused, then someone who advocates substance-causal origination must somehow escape the paradox of placing an agent at the beginning of beginningless series of causings of causings. It doesn't help to say that the agent causes the whole infinite series at once, since he must then cause the causing of it, and so forth, so that a new infinite regress begins, perpendicular to the first. It is equally paradoxical to assert that someone's causing of an act itself causes the infinity of causings of causings that divide him from the act.⁴²

In order to avoid these paradoxes, someone who advocates substance-causal origination might then either hold that causings are uncaused events, or reject the notion that causation is an event.⁴³ The position that causings are uncaused events is unsatisfactory because it seems that if causal events are to be exempted from having to be caused when all other sorts of events have to be caused, they must have some intrinsic feature that would grant them this exemption. No such special feature is evident.

A picture of substance causal origination featuring causation that *isn't* an event is equally unsatisfactory. Here the originating agent doesn't *do* anything in causing his intention or act. He is inert. His "causal relation" to his intention is like a stem growing

⁴² Roderick Chisholm embraces the notion of an infinite regression of causings of causings of an act by an agent in *Person and Object: A Metaphysical Study* (La Salle: Open Court Publishing Company, 1976), 71. According to Timothy O'Connor, Chisholm later proposed that "A causes *A's causing B* and all the other events in the nested sequence *by causing B*." (*Persons and Causes*, 58, referring to Chisholm's "Objects and Persons: Revisions and Replies", in Ernst Sosa, ed., *Essays on the Philosophy of Roderick Chisholm* [Amsterdam: Rodopi, 1979], 371-372.)

⁴³ According to O'Connor, Thomas Reid (advocating substance-causal origination) denies that causes are events. O'Connor holds that while someone's causing himself to intend to do something is an event, events of this type aren't caused. *Ibid.*, 49; 61.

out of him while he sleeps, pushing up the blossom of an intention. This unintended outgrowth of intention shouldn't give him any more moral responsibility for the intention than he would have for it if it arose randomly. In fact, this sort of origination seems to be nothing more than a random occurrence.

It might seem that in proposing that someone originates an intention to originate an intention to originate ... an (ultimate) intention, I am only taking the paradoxical proposal that an agent causes his causing of his causing of ... an intention (where the unit "of his causing" occurs an infinite number of times) and removing the agent from his initial, anchoring position. But if the agent is removed from this anchoring position, we have the fragment, "... causes his causing of his causing of ... an intention." This fragment cannot stand alone. It requires the snipped-off subject. My formulation of what goes on when someone originates an intention stands alone, however, and so it cannot be equivalent to that fragment.

If this formulation is elaborated with the causal relation inserted in the right places, it is the proposal that someone's making himself intend to make himself intend ... to do something causes him to intend to make himself intend ... to do something, which causes him to make himself intend ... to do something, which causes him to intend ... to do something. Here, "making himself intend" is to be understood as an intention-forming act which, as an act, is intended. With its infinite regress of intentions, all of which are oriented toward the ultimate act, my proposal enables us to see the originating agent as morally responsible for his ultimate intention to perform this act. The proposal that an agent causes his causing of his causing of his causing ... of an act does not permit us to

see the agent's first causing of his series of causings as an act, and so does not enable us to seem as morally responsible for his ultimate intention.

We could avoid an infinite regression if we held that while someone's origination of an intention to behave in a certain way is indeed something that he does, it is not something that he does intentionally. However, it seems to me that while our initial belief that we can originate our intentions necessarily arises in us when we are involved in a certain sort of deliberation over what to do (moral deliberation), our only enduring reason to retain this belief when we are not involved in this sort of deliberation is that we think that we must originate an act if we are to be purely moral responsible for it, and we want to be purely morally responsible for our acts. But now it seems to me that if someone's origination of an act is to purify his moral responsibility for it, then this origination would have to be something that he does intentionally. Otherwise, it would have no more moral significance than a hiccup or tremor has.

An objection that applies to any model of act-origination is that someone's origination of an intention would require that he be implausibly divided from his own mind.⁴⁴ This objection relies upon the premise that something cannot act upon itself. Someone would then have to be distinct from the subject of "his" intention if his origination act (or he himself, according to the substance-causation model) is to cause this subject to acquire this intention. This premise that something cannot act upon itself has some intuitive appeal. To the extent that it is appealing, the objection that arises from it

⁴⁴ Robert Kane's objection to the distinction between a "Kantian noumenal self" and a "phenomenal (or empirical) self that is the subject of behavior-causing "desires, beliefs, or other psychological states" is that this distinction multiplies entities (see *The Significance of Free Will* [Oxford University Press: New York 1998], 115-117). My objection to the distinction is rather than it is contradicted by introspection. I am the one who wants to be moved by certain desires rather than others, or by my evaluations rather than by my passions and urges, and so forth.

is forbidding. I cannot accept the picture of someone as distinct even from the subject of his sudden passions and urges; it is all more the difficult for me, then, to accept the picture of him as distinct from the subject of his intentions. However, the premise that something cannot act upon itself, while appealing, lacks the compelling force of, say, the premise that something cannot have contrary properties at the same time. The objection that arises from it is not fatal, then. The claim that we can originate our acts, by originating intentions to behave in certain ways, remains doubtful, but its intuitive appeal justifies our trying to see how this origination might work.

Someone would originate an intention by originating a reconfiguration of his desires. If the extent to which someone can originate a reconfiguration of his desires is limited, then it would be more accurate to say that his originative act *contributes* to the causing of his intention, as opposed to saying that he originates his intention. I will argue that someone's originative power is indeed limited. Nevertheless, it will be convenient to continue saying simply that someone originates his intention.

The simplest model of act-origination is proposed by Richard Taylor, who holds that the origination of an act simply consists in someone's being the first cause of his behavior. In causing his behavior (whether or not he is its first cause), someone necessarily behaves (that is, does what he does) purposefully, and this purpose explains his causing of it. To say that someone intends to do what he does would just be to say that he causes his doing of what he does. So, the origination of an act occurs in a single step. A second step is added by someone who, like Thomas Reid or Timothy O'Connor,

holds that in originating an act someone is the first cause of his intending to do something, which in turn causes his doing of it.⁴⁵

C.A. Campbell seems to envision a multi-step process of “absolute self-origination” that is very similar to the originative process that I propose, except that his begins with an agent as first cause, while mine begins with an infinite series of originative acts.⁴⁶ According to Campbell, “...in situations of moral temptation, ...it rests with our self absolutely to decide whether we exert the effort of will which will enable us to rise to duty, or whether we shall allow our desiring nature to take its course.”⁴⁷ In “rising to duty” (doing the right thing) we would be engaging in “purely self-originated activity.”⁴⁸ An “effort of will,” I take it, is the effort put *into* willing something; it isn’t itself the willing of something. Campbell’s references to “effortful willing” and “effortful volitions” support this reading.⁴⁹ Campbell’s will seems to be the same as my “intention”.

Campbell refers to the agent’s “determining” of an “act of decision.”⁵⁰ This, then, is the first step. The place of this act of decision in Campbell’s scheme corresponds to the place of the infinite series of originate acts in my own. Now, “we are aware of the effort as issuing from our *conscious decision to exert it*.”⁵¹ This sounds like the causing of the effort by the act of decision, a second step.

⁴⁵ According to Reid, “If the person was the cause of [the] determination of his own will, he was free in that action” (*Essays on the Active Powers of the Human Mind*, 265.) O’Connor endorses Reid’s approach, identifying Reid’s “determination of the will” with the beginning of an intention (*Persons and Causes: The Metaphysics of Free Will*, 44, 60-61).

⁴⁶ *In Defence of Free Will With Other Philosophical Essays* (London: George Allen & Unwin Ltd., 1967), 41.

⁴⁷ *Ibid.*, 42.

⁴⁸ *Ibid.*, 41.

⁴⁹ *Ibid.*, 49, 60.

⁵⁰ *Ibid.*, 44.

⁵¹ *Ibid.*, 71.

The “enabling” of a “rise to duty” by the effort doesn’t entail that the effort causes the rise to duty (the doing of what is right). However, when Campbell’s proposal that in making this effort we “introduce a new energy ... whereby we are able to act contrary to the felt balance of mere desire” is combined with his later proposal that through the introduction of this new energy “the energy of the weaker desire is ... ‘reinforced,’” we get the picture of the introduction of the new psychic energy as causing the emergence of a new psychic economy, the third step.

I would call this new psychic economy a new “configuration” of desires. Campbell’s reference to acting “contrary to the felt balance of desire” suggests that he ought to call it a new balance of desire. Campbell’s “new energy” looks very much like a desire to do whatever is right. However, Campbell won’t call it a desire, because it seems to him to be “different in kind from that energy of the desires which is the dynamic of at least most of our choices.”⁵²

The new energy’s “difference in kind” seems to consist in the combination of its unique orientation toward rightness (its tendency to cause the doing of whatever is right) and its unnatural origination in the “act of decision.” Those “higher” natural desires that may already incline us toward the right act are at best “self-regarding ... emanating from self-respect, pride, and the like;” they don’t include a desire to do what is right for the *sake* of its rightness.⁵³ It is as though, arising from the world, they still point toward the world, while the new energy, brought down from heaven, points heavenward. If there were a natural desire to do whatever is right, then the new energy would only be an artificial duplicate of this natural desire, and would have to be called “desire”. But it

⁵² Ibid., 57.

⁵³ Ibid., 60-61.

seems that in Campbell's view there is no such natural desire, so the new energy need not be called "desire".

Since the new energy is goal-oriented, and since its forcefulness is commensurate with that of our opposing, nonmoral desires, it is misleading, in my opinion, not to call it "desire". Its difference from the other desires is sufficiently noted by calling it "moral" and adding that it is of supernatural origin.

In any case, Campbell's remarks suggest that the new psychic economy (which I would call a new configuration of desires) resulting from the introduction of the new, desire-like "energy" would cause an inclination to do the right thing, in contrast to "the course towards which ... our desiring nature most strongly inclines us."⁵⁴ Campbell doesn't mention this originated inclination, but his picture of origination seems to demand it. Here we have a fourth step.

Is this new inclination itself the "will" to which Campbell refers (which I take to be what I call "intention"), or does the new inclination cause this will? The same consideration that leads me to distinguish between inclination and intention should lead Campbell to distinguish between someone's originated inclination to do the right thing and his intention to do it: someone's ability to engage in moral deliberation entails that at least in this one case he need not intend to follow his inclination. So, the fifth step would be the causing of an intention to do the right thing by this originated inclination, and doing the right thing would be the sixth step.

⁵⁴ Ibid., 57.

5. *Moral self-determinism*

I will follow Galen Strawson in calling the claim that we can originate our acts, and therefore the intentions from which these acts flow, “self-determinism”.⁵⁵ I prefer this label to “libertarianism”, despite the past association of the latter with the doctrine that we can originate our acts, because “liberty” is apparently synonymous with “freedom”, and I repudiate the claim that we must originate our acts if we are to act freely. “Originationism”, the most accurate label, is just too eccentric to employ. Nevertheless, I will continue to speak of someone’s “origination” of his act, rather than of his “self-determination” of it, because “origination” is at least a familiar verb, if not a very frequently used one. I endorse what I will call “moral” self-determinism: the claim that we can originate an intention to do something only for the sake of what we take to be the greater rightness of this behavior, as compared to what we are inclined to do. For simplicity’s sake, I will often say only that we originate, or can originate, an intention “to do the right thing,” or that we deliberate over whether or not to do the right thing. It may be assumed that when I say “the right thing”, I mean what we *take* to be the right thing. It may also be assumed that the alternative to doing the right thing is doing what we are inclined to do. It should be borne in mind that what we originate an intention to do, or deliberate over, may not, strictly speaking, be *the* right thing, or even right; it may be one of the right things that can be done, or may be more right than what we are inclined to do but less right than something else that we can do. It should also be assumed that when I speak of “doing the right thing” I mean doing something for the *sake* of its (supposed) rightness; this, I take it, is what the phrase ordinarily connotes.

⁵⁵ See Galen Strawson, *Freedom and Belief* (Clarendon Press: Oxford 1986), 26-27.

In defending moral self-determinism, I follow what I take to be C. A. Campbell's implicit argument for it. "There seems to me to be one, and only one, function of the self with respect to which the agent can even pretend to have an assurance of that absolute self-origination which is here at issue,"⁵⁶ Campbell tells us. In going on to urge that this "function of the self" consists in the formation of an intention to do what seems to be one's duty, because it is one's duty, in opposition to "the urge of my desiring nature"⁵⁷, Campbell makes clear his opinion that the agent involved in deliberation over whether to do his duty rather than what he is inclined to do not only *can* believe that he is able to originate an intention to do his duty but *must* believe this: "No amount of introspective analysis, so far as I can see, even tends to disprove that we do as a matter of fact believe, in situations of moral temptation, that it rests with our self absolutely to decide whether we exert the effort of will which will enable us to rise to duty, or whether we shall allow our desiring nature to take its course."⁵⁸ I assume Campbell would say either that what makes it our duty to do something is that this is the right thing to do, or that "X is our duty" and "X is right" are synonymous expressions.

In summarizing what I take to be Campbell's implicit argument, I adapt it to my own use by making it refer to the origination of intentions rather than to the origination of acts. This is the argument: (1) our initial belief that we can originate our intentions necessarily arises in us when we are deliberating over whether or not to do something other than what we are inclined to do, and doesn't have to arise in us when we are not involved in that sort of deliberation; (2) deliberation over whether or not to do something other than what we are inclined to do must be deliberation over whether or not to do the

⁵⁶ Ibid., 41.

⁵⁷ Ibid.

⁵⁸ Ibid., 42.

right thing, when this isn't what we are inclined to do; so, (3) our initial belief that we can originate our intentions necessarily arises in us when we are involved in this sort of deliberation, and doesn't have to arise in us when we are not involved in this sort of deliberation. But, (4) it is most reasonable to maintain that when we must believe that we can originate our intentions, we can originate them, but when we don't have to believe this we can't originate them. So, (5) it is most reasonable to maintain that we can originate intentions to do the right thing, but not any other sort of intention.

Campbell really begins with the third step of this argument, but the previous two steps seem to be implicit. The implicit first step is supported by the consideration that someone must think of his intending to do something other than what would suit his desires as an overruling of his personality. Thinking of this intention as an overruling of his personality, he cannot think of it as the effect of his personality, since he cannot think of his personality as overruling itself. But someone must think of his intention either as the effect of his personality or as originated by him. So, he must think of the intention as originated by him.

The second step of the argument is that deliberation over whether or not to do something other than what we are inclined to do must be deliberation over whether or not to do something that we think would be more right than what we are inclined to do. What supports this conclusion? We *cannot* appeal to the consideration that only an act's greater rightness could give us a reason to perform it, or explain our performance of it, since an act's greater *goodness* would yield an explanation of our performance of it that is at least as compelling. Perhaps it is more compelling, because it doesn't rely upon psychological factors, as the explanation in terms of rightness does with its evocation of a

sense of obligation. Even if we can deliberate over whether to do something other than what we are inclined to do only when our intending to do this could make us morally better, we might still deliberate over whether to behave in a way that is better, as opposed to more right, than the way in which we are inclined to behave. At least three kinds of cases in which we might deliberate over whether to behave in a better, as opposed to more right, way seem possible.

First, we might think that what we are inclined to do and what would be better are both right to the same extent. For example, someone might deliberate over whether to be what I will call “saintly”, contributing an extraordinarily large percentage of his income to charity, or over whether or not to be extraordinarily helpful, going “above and beyond the call of duty.” Second, we might completely ignore the question of rightness and wrongness, or even regard it as nonsensical, and so focus solely on the question of goodness and badness. For example, someone toying with the idea of “revolutionary commitment,” who explicitly repudiates any concern with rightness and wrongness, might deliberate over whether to slaughter masses of innocent people, which he is not inclined to do (he is a naturally sympathetic person), in order to greatly improve the general human condition. Then, there is a third, weirder case in which we might deliberate over whether to do what we think is best but *less* right than what we are inclined to do. Given that we can believe that the right act is not the best one – and the examples that I have just given show that such a belief is possible – it is clearly *conceivable* that we might be involved in such a deliberation. So, for example, a surgeon who believes that it is wrong to carve up an innocent man in order to save several other

people's lives by transplanting his organs into them might nevertheless deliberate over whether to do so for the sake of the greater goodness of this act.

Although such deliberations are abstractly conceivable, it is hard for me to imagine them as possible in the real world. It seems to me that someone who actually deliberates over whether to go above and beyond the call of duty, when he isn't inclined to, really sees this as his *personal* duty, although not the duty of the average man. It seems to me that when the revolutionary deliberates over whether to overrule his natural sympathy for individuals in order to benefit mankind in general he really thinks that the great goodness of this act makes it right that he perform it, so that his deliberation is, after all, a deliberation over whether to do the right thing. And, it seems to me that as soon as the surgeon begins deliberating, he must think that harvesting the innocent man's organs is the right thing to do. The reason that he isn't inclined to do this therefore can't be that his predominant desire is to do whatever is right. Rather, his natural sympathy for the innocent man and other nonmoral desires incline him against the act.

A rather cold-heartedly conscientious surgeon might indeed be inclined not to extract a living, innocent man's organs only because he so strongly wants to do whatever is right, and thinks that doing so would be wrong. According to my proposal, he cannot deliberate over whether to cut the man up until his desires and moral beliefs alter in either of these two ways: (1) he begins to think that it is right to cut the man up, while his continuing inclination not to do so is now caused by the new predominance of nonmoral desires (for example, to avoid harming people and conform to past practices); or (2) he becomes inclined to cut the man up because of the new predominance of perverse, nonmoral desires, while continuing to see this as wrong.

The case of the prospective saint, who supposedly deliberates over whether to go beyond the call of duty, seems more plausible than the case of the prospective revolutionary, who repudiates any concern with right and wrong, which in turn seems more plausible than that of the surgeon, who deliberates over whether to do what is less right in order to do the best thing. However, once we admit that the case of the saint is possible it seems that we have to admit that the case of the revolutionary is possible. If the better-ness of an act can be a reason to choose it over the act toward which we are inclined, even when this better-ness doesn't make it more right, there doesn't seem to be any need for an act to be thought right at all, in deliberating over whether to perform it. But if this is so, then it seems that it would be arbitrary to dismiss the possibility of a case like that of the surgeon, which is the least plausible of the three (because most difficult to imagine oneself involved in). So, if we are going to dismiss the possibility of a case like that of the surgeon, we must also dismiss the possibility of a case like that of the prospective saint.

I cannot altogether dismiss the possibility that someone might deliberate over whether to do the best thing without thinking of the act's greater goodness as making it right. Such deliberations are indeed abstractly conceivable. However, they seem unlikely. So, I endorse Campbell's implicit claim that whenever we consider not following our inclination we consider doing the right thing. (Strictly speaking, we would consider doing what we *believe* to be *more* right, for the sake of this greater rightness.)

Perhaps my impression that inclination-challenging deliberation concerns whether to do the right thing is due to my impression that the sense of *obligation* attending the thought of doing the right thing is the sense of a force opposing and neutralizing that of

desire – a force that may be internal, arising perhaps from one’s rational nature, or external, perhaps somehow emanating from a commanding divine will. We can’t feel able to do anything other than what we are inclined to do, unless we feel that we *ought* to.

I will call deliberation over whether to do the right thing “moral” deliberation. I will apply the label “moral” *only* to this sort of deliberation, even though it is also ordinarily used to indicate someone’s deliberation over how to put into effect his already-established general intention to do whatever is right. Note that if someone merely *suspects* that a possible act is more right than the one that he is inclined to perform, and is merely reviewing his possible courses of action to see whether this is so, he may be disposed to begin moral deliberation. However, he isn’t yet morally deliberating, since he isn’t in a position to form an intention to do anything other than what he is inclined to do. He cannot morally deliberate until he concludes that some possible act is indeed more right than the one toward which he is inclined. Usually, someone would bounce back and forth between morally deliberating and reconsidering the rightness of the acts in question. He would hope to discover that what he is inclined to do is right after all.

Step two must derive its support directly from introspection, then, and this takes us to step three, Campbell’s starting point. Someone thinks he can originate an intention whenever he thinks himself able to do something other than what he is inclined to do, and *only* then. He can only think this when he is deliberating over whether to do the right thing. So, he must think himself able to originate an intention whenever he deliberates over whether to do the right thing, and only then. In thinking it possible that he will decide to do the right thing, he thinks it possible that he will originate an intention to do it. (His decision would be his forming of the intention to do it. On the other hand, if his

origination of this decision is itself an act, as I hold to be the case, then the formation of his intention to originate the decision would itself be a decision.)

In agreeing with Campbell that it is *only* when we are involved in moral deliberation that we have to think that we can originate an intention, I do not deny that when our deliberation is nonmoral, and we are about equally inclined to perform each of the acts in question, we may believe that we will originate our decision to perform one of them. It is just we don't *have* to believe this. We may have the feeling that what Campbell calls our "desiring nature" is evolving toward a decision.

Galen Strawson urges that, on the contrary, whenever we must make "difficult" choices, whether or not these choices are morally significant, we must "believe that we are free" – that we can originate our decision: "And the notion of moral responsibility doesn't have to be mentioned at all. Difficult choices, one could say, are the fundamental experiential guarantors of the belief in freedom."⁵⁹ (By "freedom", Strawson means the ability to originate intentions and acts.) It seems to me, though, that unless this "difficulty" of which Strawson speaks has a morally significant character, one might feel that the difficulty can be worked out in some automatic manner by one's behind-the-scenes psychic machinery.

Richard Taylor seems to think that if we originate *any* of our acts, then surely only those that are apparently motivated by factors such as "fear, threats, a compulsive craving..." are *not* originated by us.⁶⁰ Taylor holds that someone acts when he causes his behavior, and that someone originates his act when he is the first cause of his behavior rather than being caused to cause it. He explicitly repudiates the notion that someone's

⁵⁹ *Freedom and Belief*, 70.

⁶⁰ *Action and Purpose*, 151.

desires might cause him to cause himself to behave in a way that would satisfy them; to say that someone does something because he wants X is just to indicate his goal in doing it, and not really to assert that anything causes his doing of it. Presumably, then, what permits us to conclude that we don't originate acts motivated by fear, threats, or cravings is simply that these acts don't *feel* originated by us. Taylor seems to hold that whenever it *does* seem to us that we originate our acts, it is most reasonable to think that we originate them. This position might also be taken by a self-determinist who doesn't hold, as Taylor does, that non-originative causation is itself caused. It seems to me, though, that our belief that we can originate our behavior is not enough to justify the conclusion that we can do so, given the mysteriousness of the idea of origination. This conclusion is only justified when we *have* to believe this.

Taylor urges that "there is no basis whatever" for the claim that someone's desire causes his act, "*except* the plausibility of the philosophical notion that *everything* is caused – which is ... the very thing at issue".⁶¹ Taylor thinks of action as behavior caused by the agent, so he would interpret the claim that someone's desire causes his act as the claim that someone's desire causes him to cause his behavior. When he denies that everything is caused then, he wants to deny that someone's causing of his behavior is caused. A self-determinist who, like Taylor, advocates a substance-causal account of origination (as Taylor does) but who holds, as I do, that no causation is ever caused might analogously repudiate the "philosophical notion" that every event is caused by some prior event.

The notion that every event is caused by some prior event is a very deeply rooted, psychologically compelling one, though, and so it should be repudiated only when it is

⁶¹ Ibid.

opposed by another philosophical notion that is just as deeply rooted and psychologically compelling. The notion which we have, when we are involved in moral deliberation, that we can originate a decision to do the right thing, is as deeply rooted and psychologically compelling. If this notion were incompatible with the notion that every event is caused by some prior event, then the notion that everything is caused might have to be repudiated where morally significant decisions are concerned. But the notion that we can originate a morally significant decision is *not* incompatible with the notion that all events are caused. Origination can be understood as a beginningless series of originative acts, each caused by the one preceding it, which terminates in the final decision to perform the act in question. The agent would be the subject of each one of these acts. Each act is the formation of an intention to form an intention ... to perform the act in question. Though beginningless, the causal series of originative acts would have a beginning in time, namely the moment of decision. The originative acts would be simultaneous, and simultaneous with the decision in which they terminate.

The belief in our originative power that we have when we are involved in moral deliberation is neither a theory that we entertain so as to make sense of prior evidence, nor any sort of immediate intuition of reality; rather, we simply believe this about ourselves because we have to, just as we have to believe that other people are conscious beings rather than soft robots. It may be objected that our belief that other people are conscious beings is also supported by an inference from what is evident: they behave in the way that I do, I am conscious, and so they are probably conscious. However, our belief that other people are conscious beings precedes any such supporting inference. This belief is therefore of a kind with our belief that we can originate our acts, even

though our belief that we can originate our acts is *not* supported by an inference from what is evident. We are unable to engage in moral deliberation without implicitly believing that if our deliberation is completed by a decision to do something, our deciding to do something will be our origination of an intention to do it. Here I go a step beyond Richard Taylor, who conceives of someone as originating an act by causing it without being caused to do so. Taylor only holds that the “conception” of agents as first causes of their acts “is consistent with our data, *reflecting* the presuppositions of deliberation, and appears to be the only conception that is *consistent* with them.”⁶² I hold that the moral deliberator presupposes that he can originate an act over which he is deliberating, not just that he presupposes other things that would require his being the first cause of this act.

It is one thing to accept Campbell’s proposal that our initial *belief* that we can originate our intentions necessarily arises in us when and only when we are involved in moral deliberation. It is another thing to accept his proposal that we *do* sometimes originate intentions when our moral deliberation ends in a decision to do the right thing, and that we originate intentions only in such cases. Campbell’s reason for holding that we sometimes originate an intention to do the right thing is that, since we have to believe that we can do this when we are involved in moral deliberation, the burden of proof is upon those who hold otherwise, and is not met: “Until cogent evidence to the contrary is adduced, we are entitled to put our trust in a belief which is so deeply embedded in our experience as practical beings as to be ... ineradicable from it.”⁶³ His reason for holding that this is the only situation in which we can originate intentions appears to be that

⁶² *Metaphysics* (Prentice-Hall, Inc.: Englewood Cliffs, 1963), 52; Emphasis added.

⁶³ *In Defence of Free Will*, 44.

where there is no compelling introspective evidence for our ability to originate an intention, it should be concluded that we have no such ability. Perhaps Campbell has in mind that, given the usual correlation of someone's intentions with the configuration of his desires, this correlation should be explained in terms of a cause-and-effect rule that applies to all cases where we don't have an especially strong reason to think that it doesn't apply. That he has this in mind is suggested by his later mention of "the predictability of human conduct" as a reason for preferring a self-determinism that greatly restricts the range of situations in which intentions may be originated.

Whatever else someone's moral deliberation is, it is a process that involves his being disposed to begin intending to behave in one of several ways which he thinks possible, ways that he reviews and compares. When someone who has been thinking about whether to do A or B begins to intend to do A, we say that he "decides" to do A. Someone's decision to do the right thing is his beginning to intend to do the right thing. *After* someone's decision to do the right thing, it is possible for him to think that this decision was caused by some combination of events that he did not originate. These events might have strengthened his desire to do whatever is right so much that it became dominant, giving him an inclination to do the right thing. However, while he was involved in moral deliberation he could not think this, and while he vividly recalls his moral deliberation he cannot think this. He must think of himself as able to decide to do the right thing, and in thinking of himself as able to do this he must believe that he is able to originate the intention to behave in this way.

The possibility of moral deliberation requires an exception to the rule that someone's inclination to do something causes him to intend to do it. However, it remains

the rule that someone's intention to do something is caused by his inclination to do it. (Similarly, the flapping of a bird's wings doesn't always cause it to fly, but its flight is always caused by the flapping of its wings.) Someone who decides to do what he thinks would be more right than what he was inclined to do would originate this decision. In originating this decision, he would be originating an intention to do this – his originitive act causing the onset of this intention. But, given that someone's intention to do something must be caused by his inclination, he would have to originate his intention by originating a reconfiguration of his desires. Specifically, he would originate a desire to do whatever is right, or the strengthening of such a desire. If this desire is sufficiently strong, so that it dominates the configuration of his desires, the new configuration of his desires will cause him to intend to do the right thing, by inclining him to do it.

Someone would have to originate his moral deliberation. Suppose someone is inclined to take the last piece of cake, but thinks he shouldn't. His inclination to take the last piece of cake cannot make him intend to deliberate over whether to restrain himself instead. This inclination can only make him intend to take the cake. So, if he begins to deliberate, he is doing something other than what he is inclined to do. He must originate this deliberation, then, for the sake of its rightness.

It seems necessary to hold that we can originate a decision to do the right thing without having deliberated. This will allow us to say that while someone who deliberates over whether to do the right thing must originate this deliberation, he need not have deliberated over whether to deliberate. We thereby escape an infinite regression of moral deliberations over whether to morally deliberate. I imagine someone sitting at a bar, thinking that he really ought to seriously consider going home, rather than having another

drink, as he is inclined to do. Thus, he might be involved in moral deliberation over whether to morally deliberate about whether to have another drink or go home. But it is hard to picture him involved in a prior moral deliberation over whether to deliberate over whether to deliberate about whether to have another drink or go home.

Someone is unlikely to begin morally deliberating at the exact moment that he becomes inclined to do what he doesn't think right, or starts thinking that what he is inclined to do isn't right. So, he would almost always start out intending to follow his inclination. Then he might begin moral deliberation, suspending this intention. If he decides to follow his inclination rather than do the right thing, he has re-formed his initial intention. He has not originated this intention in the positive sense of having actively brought it about. However, he might be thought of as originating it in a negative, impediment-removing sense – by ceasing to deliberate, so that he is no longer the continuing originator of his deliberation, as he has been. As soon as he stops originating his moral deliberation he stops deliberating, and as soon as he stops deliberating his inclination again makes him intend to do what he is inclined to do. It seems to me that someone might also stop morally deliberating because he stops thinking that there is something more right than what would suit these desires. His mind just wanders away to the things that he wants, and his deliberation evaporates.

I rely upon the notion of implicit belief in proposing that when we deliberate over whether to do the right thing we necessarily believe that we can originate our intention to do it. Obviously, we never actually say to ourselves, using these words, "I can originate my intention to do something". But in conversing with other people I also implicitly believe that they are conscious beings, subjects of experience, like me. I believe this

without saying to myself, “this is a conscious being....” I am not sure what the implicitness of a belief really amounts to. But I am sure that we frequently entertain implicit beliefs. We have to engage in this sort of deliberation (moral deliberation), and we have to believe, when we engage in this sort of deliberation, that we can originate our intentions. We therefore at least sometimes have to believe that we can originate them. Maybe this doesn’t constitute evidence that we in fact *do* originate them. But the necessity of this belief at least justifies an attempt to coherently envision this origination.

I identify someone’s decision to do the right thing with his beginning to intend to do it (his intention’s coming-into-being, or formation) rather than with his origination of an intention to do it, because decisions are beginnings of intention, and the origination act is not itself the beginning of an intention. The origination act *causes* the decision – that is, it causes the beginning of the intention. In this way, I maintain the symmetry of someone’s deciding to do the right thing with his other possible decision, his decision to do what he is inclined to do. If someone’s decision to do the right thing were his *origination* of an intention to do the right thing, then it would be inappropriate to give the same name, “decision”, both to this origination of an intention to do the right thing and to his renewal of his former intention to do what he is inclined to do.

According to the moral self-determinism that I endorse, our desires determine what we will choose to do (perhaps with some admixture of randomness) unless we overrule them by originating an intention to do something else. So, according to this moral self-determination, determination by desire is our “default mode.” Another sort of moral self-determinism is conceivable. It might be called “bivalent”, as opposed to the “univalent” moral self-determinism that I endorse. According to this bivalent moral self-

determinism, we would originate *either* our intention to do what we are inclined to do *or* our intention to do the right thing. Thomas Reid and Immanuel Kant both seem to endorse bivalent moral self-determinism. Bivalent moral self-determinism seems to me to be intuitively implausible, because it denies what I take to be the evident character of our desires as determining our choices unless we overrule them. Moreover, in denying this evident character of our desires, it removes us from the one situation in which we find it necessary to believe that we can originate our intentions – the situation in which we must understand our possible “overruling” of our desires as an originaive act.

The moral self-determinism that I endorse, according to which we originate only intentions to do the right thing, can be called “rightness-oriented”. I have rejected a goodness-oriented, or rightness-and/or-goodness-oriented moral self-determinism. However, the initial appeal of a goodness-oriented, or rightness-and/or-goodness-oriented, moral self-determinism is evident. If there can be a moral self-determinism, one oriented toward rightness, or toward goodness, or toward rightness and/or goodness, then it seems that there might be a pleasure-oriented, or a power-oriented, or a knowledge-oriented self-determinism, according to which someone can only originate doing what he thinks will bring him more pleasure, or power, or knowledge, than he will gain by doing what he is inclined to do. The alternative to a moral self-determinism doesn’t have to be a self-determinism without any orientation whatsoever. However, I have never encountered an endorsement of a self-determinism oriented toward a general ultimate goal (such as pleasure, power, or knowledge) other than rightness and/or goodness. Indeed, the notion of a pleasure-oriented, power-oriented, or knowledge-oriented self-determinism lacks the intuitive appeal that a moral, rightness-and/or-goodness-oriented

self-determinism possesses for many people. Moral self-determinism seems to be the only goal-oriented self-determinism that appeals to anyone, including me.

One reason why moral self-determinism is the only sort of goal-oriented self-determinism that appeals to anyone is suggested by the apparent facts about our deliberative experience. Careful consideration of these facts shows, I think, that whenever we deliberate over whether or not to do something other than what we are inclined to do we deliberate over whether or not to do the right thing. However, a less careful consideration of the same facts might suggest to us that we may also deliberate over whether or not to do the *best* thing. But we don't seem, even at first glance, to deliberate over whether to do what we are inclined to do *or* what would bring us the most pleasure, power, or knowledge, *except* when we think that it would be *right* or *best* that we do whatever would enable us to attain these things. If we deliberate over whether to do what we are inclined to do or what would bring us the most pleasure, power, or knowledge, then we think that if we were to do what would bring us the most pleasure, power, or knowledge, we would do so for the sake of the greater rightness or goodness of doing it. So, it seems that in such cases we are really only deliberating over whether to do the best or right thing.

It is *only* during our deliberation over whether to do what we are inclined to do or whether to do something else that we must believe that we are able to originate our activity. Experience strongly suggests that we deliberate over whether to do what we are inclined to do or whether to do something else only when we deliberate over whether to do the best or right thing. So, experience strongly suggests that it is only during our deliberation over whether to do the best or right thing that we must believe that we can

originate our intentions. Our belief that we can originate our intentions and resulting activity is primarily nourished by, if it does not indeed arise from, this deliberative experience. So, if there is some one general goal that we must have in originating our activity, this would have to be our doing whatever is right or best, rather than what would bring us the most pleasure, power, or knowledge. Further consideration of our deliberative experience then leads us (or, at rate, has led me) to the conclusion that this general goal would simply be our doing whatever is right; the “or best” can be eliminated as an alternate moral goal.

A second reason for the unattractiveness of an amorally goal-oriented self-determinism (according to which one may originate an intention to do what one thinks would bring one more power, pleasure, knowledge, or whatever, than one would get by following one’s inclination) is that our desire to regard people as deserving some special sort of credit or blame for their originated intentions or for their failure to originate them is our chief motive, beyond our deliberative experience, for continuing to embrace self-determinism despite its obscurity. In the next chapter, I consider how our deservingness of this special credit or blame might be understood.

6. *Pure moral responsibility*

Even though our being morally responsible for our intentions does not require our originating them or being able to originate alternate intentions, introspection suggests that we would have some special sort of moral responsibility both for intentions that we originate, *and* for those that we have, but do not originate, when we might have originated others instead. I will call intentions of both types “O-intentions”. It seems to me that someone would have this special sort of moral responsibility for his O-intentions only when they are morally better or worse than the alternate intentions would have been. For simplicity’s sake, however, I will for the most part refrain from mentioning it in what follows. At least three different approaches to defining the special sort of moral responsibility that we might have for our O-intentions are possible. I reject the first altogether, because of its reliance upon a theory of the self and its motivation that I have rejected as mistaken. I set the second aside because it is not likely to command the respect of readers who are not already committed to certain proposals of a speculative, theological character. I endorse and adopt the third.

First, it might be proposed that when we reasonably believe someone to have an O-intention, our liking or disliking of him as a person in considering his intention should be aimed at a deeper part of his being. His moral responsibility for his intention would then be a “deep” one, as opposed to the “shallow” moral responsibility which is all that someone can have for an unoriginated intention when he was not able to originate another one. The trouble with this proposal is that the distinction between “deeper” and “shallower” parts of someone’s being would have to be explicated in a way that does not separate him from the shallower part, the desiring part, as though this part weren’t truly

part of *him*. I don't see how this can satisfactorily be accomplished. It is not even satisfactory to say that while someone's desires are truly his they are not individually essential, since they can be altered without destroying his identity. This is not satisfactory because even if someone's identity is not constituted by his present array of desires, his identity seems to be uniquely *defined* by the history of his desires as they develop and mutate in a patterned, more or less coherent way. If we did not have the actual histories of desire that we have, we wouldn't particularly care about one another. So, I reject this approach.

The second approach, which I will set aside, relies upon theological speculation. Perhaps the citizens of the World to Come, no longer active and contemplating the intentions they all had during their former lives in the present world, will know only of one another's O-intentions. Only someone's moral responsibility for O-intentions will be relevant, then. A related speculation is that they will have no memory either of their own or of one another's past intentions; yet O-intentions, and these alone, will have become features of their eternal identities. These eternal moral features will be analogous to beautiful and ugly physical characteristics. (Plato has Socrates propose something like this, in *Gorgias*.⁶⁴) People should then like those among them who are morally beautiful, and to dislike those among them who are morally ugly (these will be the right attitudes for people to have toward those among them who are morally beautiful or ugly) – much as we find physically beautiful people attractive and physically ugly people repulsive. To like people in considering their moral beauty is to like them as persons, and to dislike people in considering their moral ugliness is to dislike them as persons, since someone must be a person if he is to be morally beautiful or ugly, given that moral beauty and

⁶⁴ Trans. Donald J. Zeyl, in Cooper, 791-870, 867: 525a.

ugliness are only representations of moral goodness and badness. (I am assuming that physical beauty is something real, and not just “in the eye of the beholder”; if it is real, then surely it is good, although certainly not our duty, that we respond to it favorably, for this response expresses our aesthetic acumen.)

Yet another related speculation is that God likes or dislikes people as persons only in considering their O-intentions, and will reward or punish people only for these intentions, if he rewards and punishes them at all. This would make sense, since he himself would have programmed them to have their non-O-intentions, by setting up the world in a way that would ensure their having desires of certain strengths at times when they were unable to originate intentions to do anything other than what these desires inclined them to do. It would be odd for him to like or dislike them as persons in considering intentions of theirs that he had programmed them to have.

According to this second, theological, approach, someone’s moral responsibility for his O-intention might appropriately be called an “eternal” one, as opposed to the “mundane” (or whatever) moral responsibility which is all that someone can have for a non-O-intention. Most serious thinkers will dismiss this approach as fantastic. I will therefore endorse another approach to defining the special sort of moral responsibility that we might have for an O-intention. However, those who find the theological approach congenial may employ it as an imaginative supplement to this approach. This third approach, which I endorse, is as follows. When we reasonably believe that someone’s intention was morally good or bad, not an O-intention, we should indeed like or dislike him as a person in considering it, but at the same time we should *also* subject him to the same “objective” scrutiny that we direct toward any interesting and potentially

helpful or harmful feature of the natural environment.⁶⁵ We are incapable of maintaining two parallel, sharply distinguished simultaneous attitudes toward one another, however. So, what happens is that our moral attitude toward him (our liking or disliking of him as a person) transforms into an objective one, and then the objective one transforms back into the moral one, in an unstable oscillation. On the other hand, when we reasonably believe that someone has a morally good or bad O-intention, we should take a purely moral attitude toward him in considering his intention, only liking or disliking him as a person without scrutinizing him objectively as though he were only a feature of the natural environment. The *only* attitude that we should have toward him in considering his intention is that of liking or disliking him as a person: we should *not* also take the “objective” attitude toward him in considering this intention.

Note that, strictly speaking, we should give someone pure credit or blame for what we take to be his originated intention only if we think that this intention is morally *better or worse* than his intention to do what he was inclined to do would have been; likewise, we should give him pure credit or blame for his unoriginated intention only if we think that it is morally better or worse than one that he might have originated. The greater the difference between the moral goodness or badness of his O-intention and that of the alternatives, the more pure credit and blame he deserves.

We might distinguish, then, between a “pure” moral responsibility that someone would have for an O-intention, and an “impure” moral responsibility that he would have for a non-O-intention. “Impure” moral responsibility doesn’t sound right, though – it

⁶⁵ Here again I am guided by Peter Strawson’s discussion in “Freedom and Resentment” of the possible attitudes that we can take toward other people in response to their actions. Strawson observes that we may switch from an attitude of personal engagement with someone to what he calls an “objective attitude” (190) – the same attitude that we would take toward any impersonal feature of our environment that we have to manage in a way that is convenient for us.

suggests an admixture of dirt and scandal – so I'll just call the latter “ordinary” moral responsibility. We can say that while someone who has pure moral responsibility for a morally good or bad intention deserves pure credit or blame for it, someone who has the ordinary sort of moral responsibility for his morally good or bad intention only deserves the ordinary sort of credit or blame for it.

My position that someone would *only* be able to originate intentions to do the right thing suggests the following conclusions about what his deservingness of pure credit or blame for his intentions would require.

There would be an originated conscientiousness, but no *directly* originated versions of what I have called “idealism” and “virtue”. I have given the name “idealism” to someone’s moral goodness in intending to do what is not merely right, but the best of what is right, for the sake of its being best, and the name “virtue” to someone’s moral goodness in intending to do what is not merely right, but the best of what is right, for its own sake. Idealism and virtue might be *indirectly* originated, though, in that conscientiousness is a precondition for both. Once someone has originated a general intention to do only what is right, his natural, unoriginated desire to do whatever is best, or his natural, unoriginated sympathy and helpfulness (for example), may be strong enough to incline him to perform the best of the possible right acts.

Given that a person can only originate an intention to do the right thing, it might seem that someone would always deserve pure credit for his originated intention. However, this isn't so. I have been using the expression “doing the right thing” to indicate someone’s doing what *seems* to him to be more right than what he was inclined to do. Someone’s belief that an act is more right than what he is inclined to do may be

unreasonable. Suppose, then, that the only reasonable belief that can be held about the moral value of the act is that it is *less* right than what he is inclined to do. If he originates an intention to perform it for the sake of its rightness, this intention would be morally *worse* than his intention to do what he was inclined to do would have been. If his unreasonable belief in the act's rightness is the result of past intentions for which he deserves pure blame, then perhaps he deserves additional pure blame for this new intention as well. Someone would deserve pure credit only for his originated intention to do what he *reasonably* believes to be more right than what he was inclined to do. Only when he fails to originate such an intention, though he might have done so, would he deserve pure blame for intending to do what he is inclined to do.

When beliefs regarding rightness and wrongness are unreasonable, their correctness is accidental. Someone who correctly but unreasonably believes that he shouldn't follow his inclination might just as easily have come to believe that what he is inclined to do is right. It seems, then, that someone who correctly but unreasonably believes that act A is more right than act B, when he is inclined toward B but can originate an intention to do A, no more deserves pure blame for intending to do B than he would deserve pure credit for originating an intention to do A.

If we can only originate intentions to do the right thing, then various scenarios in which we might otherwise have deserved pure blame for originating an act would be ruled out. According to such scenarios, we would deserve pure blame for originating an intention to do something purely for the sake of its greater *goodness*, while unreasonably believing the act to be at least as right as the one that we were inclined to perform, or while unreasonably believing that right and wrong are superstitions, or even while

believing the act to be more *wrong* than the one that we were inclined to perform. In ruling out such scenarios, we rule out the possibility that revolutionaries devoid of any sense of duty would originate massacres solely in order to benefit mankind, or that they would in doing so overrule their scruples. In originating such acts, they would only be able to overrule their non-moral desires to refrain from harming others, to conform to established patterns of conduct, and so forth. Of course, a scenario in which someone would deserve the most intense pure blame for originating an intention to do something that he believes to be more wrong than what he is inclined to do for the *sake* of this greater wrongness, or for the sake of its badness, would be ruled out as well. No Miltonic Satan would be able to originate the evil that is his good.

Someone who doesn't deserve pure credit or blame for an intention or act that he originates or fails to originate when he could have might still deserve ordinary credit or blame for it. For example, Huck Finn is inclined to help Jim to escape slavery, but he unreasonably believes that doing so would be wrong. He decides to help Jim escape, but could have originated an intention to report Jim to the authorities, for the sake of the greater rightness of doing so. Surely, he doesn't deserve pure blame for helping Jim escape. In fact, most of us would be inclined to think that he deserves credit for doing so, because his ultimate goal in helping Jim to escape is something that is in fact good – Jim's being free, or happy, or whatever. The credit that he deserves would have to be the ordinary sort of credit. But now suppose he decides to help Jim in order to ignite a bloody civil war (a bad thing). In this case, although he doesn't deserve pure blame for helping Jim instead of originating an intention to report him to the authorities, he seems to deserve the ordinary sort of blame.

How then might we gratify our desire to regard conscientious tyrants and their conscientious agents and supporters as deserving pure blame for their murderous activity? It seems that under certain circumstances someone who originates wrong acts for the sake of their supposed rightness might deserve pure blame for them despite the unreasonableness of his belief that they are right. Similarly, a naturally conscientious or idealistic person whose wrong acts result from his overwhelmingly powerful unoriginated desire to do whatever is right and/or best might deserve pure blame for them. Someone's false and unreasonable belief that his murderous activity is right might depend upon some past activity of his which he at least sometimes correctly and reasonably believed to be less right than some other activity that he might have originated. Or, his belief might depend upon a present activity of this sort. For example, perhaps the young secret policeman would not have believed that the assassination of dissidents is right if he hadn't been getting drunk with zealots. Alternately (or in addition) his belief that his murderousness is right might depend upon his failure to seriously consider whether or not his project is truly right at key moments when he correctly and reasonably believed or even continues to believe that he might and *should* engage in this reflection, despite his inclination not to. If he had originated this serious consideration, or were to do so now (no longer suppressing his recurrent doubts), perhaps he would have abandoned, or would still abandon, his unreasonable belief in his project's rightness. He *might* therefore deserve pure blame for his behavior, after all.

We can't ever be sure that there have been, or continue to be, circumstances of this sort, so that the conscientious and/or idealistic wrong-doer deserves the pure blame that we would like to give him. It is possible that he never had nagging doubts, and never

even suspected, at the beginning of his career, that getting drunk with zealots, or acts with a similar tendency to distort moral perception, might be wrong. But it wouldn't be absolutely outrageous to assume these things about him.

Perhaps this sort of thing is all that anyone ever deserves pure blame for: hanging out with the wrong crowd despite one's quiet intuition that one shouldn't, ignoring one's nagging doubts, and so on. If so, then the conscientious or idealistic tyrant would at least turn out not to deserve *less* pure blame than the ordinary murderer who ignores his doubts as he plunges into the act. Anyone who is annoyed by the thought that the conscientious or idealistic tyrant would not deserve *more* pure blame than the ordinary murderer may embrace some version of the theological approach to defining the special sort of moral responsibility that someone can have for an act if he originates it or might have originated another. We are free to employ the theological approach, relying as it does upon the notion of what I have called "eternal" moral responsibility, as a supplement to (not a replacement for) the more naturalistic approach that I have endorsed, an approach relying instead upon the notion of "pure" moral responsibility. Once the theological approach is embraced, anything goes.

We can then entertain the thought that there are grades of eternal moral responsibility – of deservingness of eternal blame or praise. We might distinguish, for example, between the following kinds of cases, where the second kind of case features someone who deserves a more intense eternal blame for his morally bad intention or act than the first kind does. (By "someone believes that he ought to X", I mean that he believes that X is the right thing for him to do.) In cases of the first kind, (1) someone formerly believed that he ought to take steps to alter, or guard against the expression of,

certain aspects of his personality (such as a mercurial temper) that he then believed might result in his behaving in this way, and/or he formerly believed that he ought to carefully assess the accuracy of beliefs which he suspected might result in his behaving in this way (such as the belief that certain people are malevolent); (2) this process, had he originated it, might have resulted in his now intending to behave in a way that isn't morally bad; and (3) he did not originate it, although he might have. (We can safely assume that in every case in which a normal adult has a morally bad intention, these requirements have been met.)

In cases of the second kind, where the moral responsibility for behaving badly seems to be more intense, 1) the person *presently*, while intending to behave in the way in question, believes that this way of behaving *might* be wrong (note that I am not suggesting that anyone can believe that the way in which he wants to behave is *bad*); (2) he presently believes that he ought to carefully consider the moral value of this way of behaving in order to make sure that it isn't wrong; and (3) he doesn't originate this careful consideration, although he might. (Note that if we replace the "originate" in these cases with a simple "undertake", we get a differentiation of grades of *ordinary* moral responsibility. Thus we distinguish between cases in which we should like or dislike someone as a person to a greater or lesser extent in considering his intention, even where our liking or disliking of him as a person (our "moral attitude" toward him) is rightly mingled with an objective, managerial attitude toward him.)

I have proposed that someone who originates an act does so with an *originative* act that he in turn originates, and so on forever. It might be urged, then, that we think of the extraordinary significance of someone's moral responsibility for his originated act as

consisting merely in his being morally responsible in the ordinary way for every “inner” originaive act in the series, as well as for the ultimate, “outer” act. Quantity – the infinite multiplication of instances of ordinary moral responsibility – would perform the function that difference in quality does in my own proposal regarding the “purity” of someone’s moral responsibility for his originated acts. One difficulty with this suggestion is that it seems to entail, bizarrely, that in cases in which someone doesn’t originate his act, but might have originated another act for the sake of the greater rightness that he reasonably attributes to it, he would be guilty of an infinite quantity of sins of omission, corresponding to the infinity of originaive acts that he might have performed but hasn’t. Moreover, this alternate proposal depends upon an account of origination (as consisting in an infinite series of originaive acts) that is not intuitively compelling. Since this account of origination is not intuitively compelling (although I do believe it to be the best possible account of origination), it cannot explain what I take to be the intuitive appeal of the notion that the power to originate acts would necessarily be accompanied by an extraordinary moral responsibility.

Finally, it doesn’t seem that we like or dislike someone as a person in considering what we take to be his purely inward act of origination, or his omission of such a purely inward act. Only someone’s ultimate, outer act, or his intention to perform such an act, concerns us, and can move us to like or dislike him as a person in considering it. In assuming a purely moral attitude toward someone in response to what we take to be his O-intention, we continue to focus upon that intention, and not upon his performance of the prior acts in the originaive series terminating in this intention. His origination of the intention, or his ability to originate another, is just significant background information.

Its significance seems to be that it ensures the purely moral character of our attitude toward him, not permitting this attitude to dissolve into an objective, managerial one.

We can distinguish between a *positive* pure moral responsibility for an originated intention, and a *negative* pure moral responsibility for an unoriginated one, when another might have been originated. I have argued that someone deserves credit only for his intentions and acts, since his acts, understood as “greater acts”, include within themselves any generally intended consequences of his core acts. However, someone deserves blame for any foreseeable bad consequences of his acts, even if he did not intend to bring them about by means of his core act. But I have also argued that someone may deserve pure blame even for his originated intention, when his unreasonable belief in the intended deed’s rightness is due to decisions of his for which he deserves pure blame. (It doesn’t seem that someone can deserve pure credit for an *unoriginated* O-intention – one that he has when he can originate another.) So, someone can have positive as well as negative pure moral responsibility not only for his O-intentions and O-intended acts, but also for these acts’ foreseeable bad consequences.

It is not enough to say that someone has positive pure moral responsibility for his *behavior* if he originates it, or negative pure moral responsibility for his behavior if he does not originate it but might have originated other behavior. Someone seems to be morally responsible only for his *acts* – his intentional behavior – and his intentions to behave in certain ways. So, someone’s direct origination of his behavior – which is what Richard Taylor pictures us doing⁶⁶ – wouldn’t be enough to give him positive pure moral responsibility for this behavior. Nor would his ability to directly originate alternate behavior give him negative pure moral responsibility for acts that he doesn’t originate. If

⁶⁶ See *Action and Purpose*, 111-112 (“Agents as Causes”).

he is to have positive pure moral responsibility for his behavior, he must originate this behavior by originating the intention that causes it. And, if he is to have negative pure responsibility for his behavior, he must be able to originate an intention to behave in another way. Perhaps Taylor would say that the origination of behavior is itself a way of intending it. It seems to me, though, that this origination would have so little in common with non-origination that no substantive resemblance would be indicated by the application of the name “intending” to both. If non-origination makes the intended behavior an act for which someone can be morally responsible, origination cannot do so, since it is something else entirely.

When someone doesn't originate his acts, something other than he himself must cause him to have the desires that together cause him to intend to behave as he does, or must at any rate cause him to be the sort of person that he is, tending to desire what he desires as strongly as he does, so that he intends to do what he intends to do. We might say, then, that he “naturally” intends to behave as he does, and that his behavior is caused by his naturally intending to behave in that way.

It might seem that a third sort of pure moral responsibility (in addition to positive and negative pure moral responsibility) should be incurred by someone who apparently *plunges* into the behavior that he naturally intends to perform; such a person seems to consciously repudiate the possibility of behaving in another way. Thus, someone who has been struggling against the inclination to have a third beer suddenly thinks, “The hell with it, I'm going to finish the six-pack,” and with a sort of grim exultation heads for the refrigerator. But we can understand this kind of thing as a case either of positive pure moral responsibility or of negative pure moral responsibility. We can understand the

sense of plunging as being a sort of involuntary emotional manifestation of the sudden stopping of intense deliberation, in which case his pure moral responsibility for it remains merely negative. Or, we can understand it as an instance of someone's making himself intend a more intense performance of the behavior that he had previously intended, in which case he incurs positive pure moral responsibility for a corresponding increase in the intensity of the intended behavior. So, for example, someone who has been suffering from a guilty conscience over his daily consumption of three beers, which he had naturally intended, perversely originates an intention to drink a six-pack a day; he now has positive pure moral responsibility for half of his daily drinking. I hold that someone can have positive pure moral responsibility for an intention only when he intends to do the right thing. So, the beer-drinker with positive pure moral responsibility for his six-pack-a-day decision must have come to believe that drinking that much beer every day is right, perhaps because of the goodness of being in a beery state of mind.

I hold that the origination of an intention is itself an originated act. It follows, then, that someone who has positive pure moral responsibility for his outer act, and for his intention to perform it, has positive pure moral responsibility for this origination act, and for his intention to perform it.

At the end of the previous chapter, I offered one explanation of why it is that a self-determinism oriented toward a general ultimate goal other than rightness or goodness, such as pleasure, power, or knowledge (pursuit of which would be an alternative to following our inclination) doesn't appeal to us. The explanation that I offered there was that our deliberative experience doesn't support an amorally goal-oriented self-determinism. Our gut feeling that we would deserve a special sort of credit

or blame for originating an intention to do the right thing or the best thing suggests an additional explanation for the unattractiveness of a pleasure-, power-, or knowledge-oriented self-determinism. An amorally goal-oriented self-determinism would make it impossible for somebody to deserve some sort of special credit or blame for his originated intentions or for his failure to originate one when he might have done so. Yet it seems that our desire to regard people as deserving some such special credit or blame is our chief motive, beyond our deliberative experience, for continuing to embrace self-determinism despite its obscurity.

I have recommended that we understand this special credit or blame as “pure”, rather than “deep” or “eternal”. Clearly, someone cannot deserve credit for an originated intention to do something for the sake of the pleasure, power, or knowledge he hopes to enjoy or to acquire in or through doing it, if the only reasonable view is that it is right or good that he follow his inclination instead. So, if he is to deserve pure credit for an originated intention to do something for the sake of pleasure, power, or knowledge, the intended behavior must be more right or better than what he is inclined to do. But either (1) it is merely accidental that his originated intention is an intention to do what is more right or better than what he is inclined to do, or (2) he originates this intention for the sake of the greater rightness of the intended behavior, or (3) he originates it for the sake of the greater goodness of the pleasure, power, or knowledge that he expects to enjoy or to acquire through it. If it is merely accidental that his originated intention is an intention to do what is more right or better than what he is inclined to do, then he cannot deserve pure credit for it. So, he deserves pure credit for his originated intention only if he originates it for the sake of the greater rightness of the intended behavior, or for the sake of the

greater goodness of the pleasure, power, or knowledge that he expects to enjoy or to acquire through it. But then we have a self-determinism oriented toward rightness and goodness after all.

I have argued that careful consideration of our deliberative experience leads to the conclusion that when we deliberate over whether or not to do something other than what we are inclined to do, we focus upon the greater rightness, not the greater goodness, of this alternative to following our inclination – except insofar as it is clearly its greater goodness than makes it more right. So, it is reasonable to conclude that someone would deserve pure credit for an intention to do the *right* thing, not the best thing.

7. *Moderate moral self-determinism and behavioral probability*

According to the moral self-determinism that I endorse, the intentions that we can originate must be intentions to do what we think is more right than what we are inclined to do. We would intend to do it at least in part for the sake of its greater rightness. Originated intentions need not be intentions to do what is thought *most* right. They might also be “compromise” intentions to perform acts toward which we are less inclined than we are toward others that seem less right, but toward which we are more inclined than we are toward others that seem more right. Nevertheless, since it is convenient to call all intentions that aim at rightness intentions to “do the right thing”, I say that originated intentions in general are intentions to do the right thing.

I have rejected an alternate sort of moral self-determinism according to which someone might originate an intention to do something for the sake of its being *better* than what he is inclined to do. This sort of moral self-determinism might be represented as the position that originated intentions must be intentions to do the *best* thing. However, anything that I have to say about the “rightness-oriented” moral self-determinism which I endorse can easily be adapted to fit a goodness-oriented (or rightness and/or goodness-oriented) version of moral self-determinism.

“Amoral” self-determinism wouldn’t require that our originated intentions be intentions to do the right or best thing. It might be amorally goal-specific, requiring that our originated intentions be intentions to do the most pleasant, empowering, or knowledge-enhancing thing. However, this amoral goal-specificity is unappealing, as we have seen. Proponents of amoral self-determinism are likely to deny that our originated intentions must be of any one type.

We can now distinguish between “extreme” self-determinism and “moderate” self-determinism. Both might be moral or amoral. According to *extreme* self-determinism, the configuration of someone’s desires cannot make it impossible that he originate an intention to behave in a certain way. Someone can originate an intention to do anything he thinks his skills and environment will permit. If he is mistaken in thinking that his skills and environment will permit him to behave in the way in question, then while he may originate an intention to behave in that way he will of course not be able to originate behaving in that way. According to extreme self-determinism, someone can originate any behavior for which he has the requisite skills, and which his environment makes possible.

Extreme *moral* self-determinism would assert that another factor can make it impossible for someone to originate his behaving in a certain way. If someone is to originate an intention to behave in a certain way, he must think that his behaving in this way is more right than what he is inclined to do, because he can only originate an intention to do the right thing. So, if he does not think that a certain way of behaving is more right than what he is inclined to do, he cannot originate an intention to behave in that way. According to extreme moral self-determinism, then, as long as someone’s environment and skills permit his behaving in a certain way and he thinks that this behavior is more right than what he is inclined to do, he can originate his behaving in this way. He can originate an intention to do anything that he thinks is more right than what he is inclined to do, when he thinks his skills and environment will permit him to do it.

According to moderate self-determinism, on the other hand, the configuration of someone’s desires *can* make it impossible that he originate an intention to behave in a

certain way. So, even when his environment and skills permit his behaving in this way, the configuration of his desires might make it impossible that he originate his behaving in this way. Someone's personality – his disposition to have certain desires of certain strengths in certain circumstances – can therefore make it impossible that he behave, or merely intend to behave, in a certain way. A very irritable person, for example, will not be able to respond politely to a provocation. A moderate *moral* self-determinism would assert that someone is not only unable to originate an intention to do what he doesn't think more right than what he is inclined to do. He may even be unable to originate an intention to do the right thing, despite his having the skills required for it and despite the absence of environmental impediments.

I have rejected a "bivalent" moral self-determinism according to which we would originate *either* an intention to follow inclination *or* an intention to do the right thing. Bivalent moral self-determinism would have to be extreme. Given that someone would originate either an intention to follow inclination or an intention to do the right thing, his ability to originate an intention to do the right thing could not be limited by the strengths of his desires. His desires would have strengths only in relation to each other. Whenever someone thinks that his behaving in a certain way would be more right than doing what he is inclined to do, his desires would have no actual strength until he has originated an intention to follow his inclination.

It seems to me that according to *any* self-determinism, the causing of someone's final, originated act by his final *originative* act must consist in his making himself so intensely desire something new, or in his making himself desire something old so much more intensely, so that the configuration of his desires is shifted in favor of this new, or

newly intense, desire. This will then cause him to intend to do something that will be or involve the bringing-about of the object of this desire to a greater extent than would otherwise have been the case. Someone's ability to originate an act therefore consists in an ability to generate a new desire, or to increase the intensity of an old one. According to the moral self-determinism that I endorse, the relevant desire would be a desire to do whatever is right.

"Willpower" seems to be an appropriate name for this ability to make oneself intend to do what one isn't inclined to do by generating a new desire or by increasing the intensity of an old one. The dispute between moral and amoral self-determinism, then, is over whether someone's willpower is only his ability to make himself intend to do the right thing, or his ability to make himself intend to do whatever seems possible to him. The dispute between extreme and moderate moral self-determinism is a dispute over whether or not someone always has enough willpower to make himself intend do the right thing, when he thinks that his skills and environment permit it.

In proposing "willpower" as an appropriate name for someone's ability to make himself intend to do something other than what he is inclined to do, I am supposing that we ordinarily use this word to refer to such an ability. However, I do not deny that we also use the same word to denote someone's disposition to forgo the immediate gratification of some of his desires in order to make progress toward achieving more distant goals. According to this use of the word, which may in fact be more ordinary than the one that I have put it to, someone's having a lot of willpower simply consists in the fact that he is often inclined to do what he doesn't want to do. Great willpower might be attributed to someone who is disposed to climb a steep hill in order to get to a party.

Non-self-determinists may use the word “willpower” in this way, as indicating someone’s ability to do things that he doesn’t want to do. For example, Hare invokes insufficient willpower as the solution to the problem of weakness of will: the weak-willed person, he says, is one who is physically able to “resist temptation”, but not psychologically able. That he nevertheless has the intention that makes a person morally good is indicated by the weak-willed person’s “remorse”, for, “though unable to overcome this temptation, he keep[s] alive the will-power which may overcome lesser ones”.⁶⁷ The weak-willed person simply does not have enough willpower to overcome his desires. Hare’s discussion of the compatibility of moral responsibility with determinism shows that he does not think of willpower as an ability to resist temptation specifically through the origination of acts other than those toward which one is tempted.⁶⁸

I share Hare’s opinion that moral responsibility is compatible with determinism. However, unlike me, Hare doesn’t show any concern that there might be a special sort of moral responsibility for an act which would require the origination of this act, or the ability to have originated another act. So, since Hare embraces a notion of willpower without embracing the notion that acts can be originated, Hare’s willpower is evidently not an originaive ability. Similarly, George Ainslie applies the name “willpower” to a psychic law governing the rate at which the intensity of desire for certain things decreases when the attainment of these things is consigned to a more and more distant future.⁶⁹

It is tempting to say that willpower is simply potential desire – whether potential desire for anything at all, according to amoral self-determinism, or potential desire for

⁶⁷ *Freedom and Reason* (Oxford: Oxford University Press, 1963, 80.

⁶⁸ *Ibid.*, 63-66.

⁶⁹ See his *Breakdown of Will* (Cambridge: Cambridge University Press 2001).

whatever is right, according to moral self-determinism. This way of talking is misleading, however, since it suggests that willpower is a sort of entity crouched within an agent's psyche, waiting to be transformed into another sort of entity – desire. I have rejected this way of thinking about desire; rather it is only a state or continuing event of which the desiring person is the subject. Similarly, willpower is only an ability – someone's ability to originate desire – not any sort of entity. The state or continuing event of “idle fancy”, or wishing, typically expressed by thoughts of the type, “Wouldn't it be nice if...”, might perhaps be the prior seed of an originated desire. If so, then wishing is potential desire. However, wishing obviously isn't willpower; nobody would be inclined to think that the two are the same. Rather, wishing would be converted into desire (if that is what happens) *by* the deployment of willpower.

Another misleading image is that of willpower as a sort of extra fuel-supply for a motivational engine whose operation is desire. Here desire is not imagined as an entity inhabiting a person or pushing him around from the outside. Desire is only the *operation* of the motivational engine – either a continuing event, or state of being “on”. An infusion of willpower into the engine makes it work harder, intensifying desire. However, since it is now the motivational *engine* that inhabits someone, pushing him around from within as though he were a car, this image really isn't an improvement on the one that places desire itself in this role. Someone's act still turns out to be performed by something other than himself. This is the first and most important objection to this image. A second objection to it is that when the image of desire as the operation of a motivational engine is incorporated into a full-scale psychic model, it becomes necessary to envision a bizarrely complex, inefficiently designed vehicle. Every desire that

someone has would have to be the operation of a distinct engine that pushes him in a certain direction. Or at least, this would have to be true of someone's basic desires. He would have a pleasure-engine pushing him toward pleasure, a power-engine pushing him toward power, a knowledge-engine pushing him toward knowledge, a rightness-engine pushing him toward rightness, and, perhaps, a distinct goodness-engine pushing him toward goodness. Where he goes would be determined by how powerful these various engines are. According to the (rightness-oriented rather than goodness-oriented) moral self-determinist application of this image, the willpower-fuel could only be channeled into the rightness-desiring engine. According to an amoral self-determinist application of this image, on the other hand, someone can channel his willpower-fuel into any of these engines, as though by turning a psychic steering wheel. According to an extreme self-determinist application of this image, there would be an infinite quantity of willpower-fuel, enabling someone to so intensify the operation of the motivational engine into which he pumps it that this motivational engine can prevail over all of the engines that push him in other directions. According to a moderate self-determinist application of this image, there would be a limited quantity of willpower-fuel. The craziness of such a vehicle, as an engineering project, suggests that this is not at all how motivation works. A third objection to this image is that while it doesn't portray desire as a substance (portraying instead as the operation of a substance, namely the motivational-engine), it still portrays willpower as a substance (for gasoline, after all, is a substance), and this is just as bad.

The safest thing to do, then, is simply to stick with the conception of willpower as an ability to originate desire, while remaining on guard against the misleading influence of images that inevitably arise as illustrations of this conception. I admit that not much is

clarified by saying that willpower is an “ability.” What, after all, is an “ability”? It won’t do to imagine a sort of backpack-generator connected to a power-tool, or anything of that sort. I suppose I can plead that the notion of ability is “primitive”.

For one person to have *more* willpower than another person has, then, would be for him to have an ability to originate a stronger desire. (That is, to have more ability to do something is to have the ability to do it more intensely. Someone who has a greater ability to lift weights than another person has is able to lift more weight, or to exert a greater lifting-force, than the other can.)

It is convenient to say that someone who originates his act, or his intention to behave in a certain way, or his desire or the strengthening of his desire for something, “makes himself” act or intend to behave in that way, or want or more strongly want the thing in question. The reader should bear in mind that whenever I speak of someone’s “making” himself act, intend, or want I mean that he originates his act, intention, or desire. Obviously, we often speak of someone’s making himself want or intend to do something without having any thought of his having originated his desire or intention. According to this ordinary use of the expression, someone might be said to make himself want to take a daily early-morning run by intentionally doing so, day after day, even if his intending to do so is caused by a desire for health whose strength in him is caused by his heredity and past experiences. However, if I say here that someone makes himself want to take a daily early-morning run, I will mean that his heredity and past experiences are not sufficient causes of his intention; that without his origination of at least some part of the strengthening of his desire for health he would not have intended to run day after day, and so would not have come to want to do so.

Someone might make himself intend to behave in a certain way, thereby incurring positive pure moral responsibility for this behavior, either by originating a new desire or by strengthening an old one. Let us consider a simplified moderate moral self-determinist scheme under which someone can only desire pleasure or whatever is right. Suppose that his only present desire is for pleasure, but that at the same time he is able to originate in himself a desire for whatever is right that is stronger than his present desire for pleasure. Now, suppose that he has three behavioral options. He can eat a candy-bar, a banana, or a vitamin-fortified “energy-bar”. The energy-bar is not as pleasant as the candy-bar but more pleasant than the banana, healthier than the candy-bar but less healthy than the banana. He thinks that the healthier a food is for him, the more right it will be that he eat it. Eating the banana will be *the* right thing for him to do, but eating the energy-bar will still be more right than eating the candy-bar. Suppose that he can make himself intend to eat the banana – he has enough willpower for this. Still, assuming that he doesn’t have to deploy *all* of his willpower, he might only make himself want whatever is right just as much, but no more, than he wants pleasure. In this case, then, he would make himself intend to eat the energy-bar. Now suppose that he only has enough willpower to make himself want whatever is right as strongly as he wants pleasure. In this case, he can only make himself intend to eat the energy-bar; he can’t make himself intend to eat the banana. Perhaps, though, he begins with a desire for whatever is right that is weaker than his desire for pleasure, and so begins by intending to eat the candy-bar. He suspends this intention by entering into moral deliberation. Now, if he has enough willpower to strengthen his desire for whatever is right so that it is stronger than his desire for pleasure, he may in this way make himself intend to eat the banana. If he

only has enough willpower to strengthen his desire for whatever is right to the point of parity with his desire for pleasure, however, he will only be able to make himself intend to eat the energy-bar.

According to an extreme moral self-determinism, he would always be able to make himself intend to eat the banana, so that eating the banana would be a live possibility, a viable option, for him, and not just something that is possible for him in the sense that it is physically and logically possible and he has the required skills. According to moderate moral self-determinism, it might not be a live possibility for him to eat the banana. It would be a possibility in a certain sense – he would have the required skills – and so he might deserve the ordinary sort of blame for not doing it, but since his personality removes it from the realm of viable options for him, he wouldn't deserve any pure blame for eating the energy-bar or perhaps even the candy-bar instead (if he doesn't even have enough willpower to make himself eat the energy-bar).

Self-determinism can be extreme and moral, extreme and amoral, moderate and moral, or moderate and amoral. The self-determinism that I endorse is the moderate, moral kind. In introducing the notion of willpower as the key to making sense of a moderate self-determinism, I am really only stressing that someone's origination of an act requires that he alter his desires. Someone's willpower is only his ability to make himself begin to desire something that he hasn't been desiring, or to strengthen one or more of his desires to a certain extent. According to the moderate *moral* self-determinism that I endorse, we are only able to generate or to strengthen a desire for whatever is right. How much this accomplishes depends upon the strength of our contrary desires for things that we think are less than fully right – both our pre-existing desires for such things, and

desires for such things that we are likely to begin to exhibit before we are able to fulfill our originated intention, because of our personalities.

Extreme self-determinism requires that we attribute real probabilities to people's originating acts of one sort rather than acts of another sort. Moderate self-determinism doesn't require this. Extreme self-determinism thus has a "probability problem" that moderate self-determinism doesn't have. This makes moderate self-determinism preferable to extreme self-determinism.

We frequently call events, including acts, more or less probable (or likely). In doing so, we either intend to attribute probabilities to these events, or we report or express the degree of our confidence that they will occur. Now, our being more or less confident that events, including acts, will occur can surely be justified. But if our partial confidence that an event will occur is justified by anything other than our recognition that it really has a certain probability, then this justification would have to consist in our recognition of the truth of determinism (the proposal that events are either necessitated by prior events, or rendered impossible by them) together with our awareness that we only have a certain degree of knowledge of the events that might necessitate or exclude the one in question. We aren't completely confident that someone will perform a certain act, because we don't know what he may have begun to desire or believe. Self-determinism is incompatible with determinism, though. If someone originates his act, then his act can't have been determined by events other than his previous originaive acts. So, self-determinists have to hold that events really have probabilities (whatever these may be), and that our utterances regarding the probabilities of events, including acts, are statements that can be true or false. Extreme self-determinists hold that we originate all

of our acts, or are always able to originate acts other than the ones that we perform. So, extreme self-determinists have to hold that our originated acts have probabilities.

Moderate self-determinists, on the other hand, may hold that while many of our acts do have probabilities, our originated ones, or the ones that we perform when we might have originated others, don't have probabilities – that they are neither necessary, nor more likely or unlikely than not to be performed.

When we think that there is a probability that an individual person will behave in a certain way, we infer this probability from what I will call the “behavioral profile” of the group of people who have natural backgrounds like him. A group's behavioral profile is a list of the percentages of its members who behave in certain ways, where these percentages (the group's “behavioral *description*”) reappear with decreasing exactness in randomly-selected portions of this group as these randomly-selected subgroups get smaller and small. The rate at which the conformity of the randomly-selected subgroup's behavioral description to that of the total group diminishes as the subgroup gets smaller and smaller depends upon how many features belong to the natural background that defines the total group. The more features belong to this defining natural background, the more slowly the conformity of the randomly-selected subgroup's behavioral description to that of the total group diminishes. If the only defining feature is having been frequently beaten as a child, then the local group has to be pretty large if it is to exhibit a behavioral description (say, that ten percent of its members commit violent crimes as adults) similar to that of the larger group. But if the defining natural background also includes having been chronically neglected, hungry, and subjected to daily mockery at school, then the group doesn't have to be as large. The

steady diminishing of the conformity of the randomly-selected subgroup's behavioral description to that of the total group continues all the way down to the point at which it is no longer possible to talk about percentages of the subgroup's members behaving in this or that way. We attribute to the individual members of the total group a probability of behaving in a way belonging to the total group's behavior profile that is equal to the percentage of people within the total group that behaves in this way. So, if 30% of the total group's members behave in this way (and this way of behaving belongs to the total group's behavioral *profile*, not just to its behavioral description), then we infer that for each individual member there is a 30% likelihood that he will behave in this way.

We make this inference because we take it that group behavioral profiles are *caused* by probabilities that individual people with certain natural backgrounds will behave in certain ways. This isn't the only way of explaining the fact that there are group behavioral profiles. Group behavioral profiles might also be caused by the fact that within a group there is necessarily a sub-group whose individuals simply have to behave in a certain way. But extreme self-determinists can't hold that individuals ever simply have to behave in a certain way. They can't hold this, because only the personalities of these individuals could necessitate their behaving in a certain way. Extreme self-determinists cannot allow that our personalities make it necessary that we behave as we do, for if this were so we would not be originators of our behavior. Therefore, extreme self-determinists can only explain group behavioral profiles as being caused by the *probability* that people with certain natural backgrounds will behave in certain ways. Otherwise, they will just have to leave the fact that there are group behavioral profiles unexplained, merely urging that this unexplained fact is compatible with the individual

group-members' origination of much of their behavior.⁷⁰ But in explaining group behavioral profiles as being caused by the probability that people with certain natural backgrounds will behave in certain ways extreme self-determinists don't make any real progress, because they still can't explain how any sort of probability that someone would behave in a certain way can be compatible with his originating his behavior.

We also can't explain how any given probability can be attributed to a die's landing on a certain face. Why should we think that it is any worse to attribute a probability to someone's originating a certain act? It seems to me to be *merely* inexplicable that a normal die should have a likelihood of 1 in 6 that it land on a certain face. It seems to me to be *impossible* that someone's origination of an act of keeping a promise or assassinating an enemy should have a likelihood of 1 in 6. It is difficult for me to justify my feeling that this is impossible rather than merely inexplicable, which is unfortunate, since it is this feeling that really motivates my endorsement of moderate, as opposed to extreme, self-determinism. The closest I can come to justifying it is to offer, as a premise, that a probability can attach to an individual event only if nothing causally necessitates it. Nothing causally necessitates a die's landing on a given face, but something *does* causally necessitate someone's origination of an act. According to my event-causal account of the origination of acts, the preceding infinite chain of his originative acts necessitates the final one. According to the more common agent-causal model of origination, the agent himself causally necessitates his origination of the act. Because originations of acts are causally necessitated, then, probabilities cannot attach to them; they cannot be more or less likely to occur.

⁷⁰ As does R.L. Franklin, who calls group behavioral profiles "macro-patterns" (*Freewill and Determinism: A Study of Rival Conceptions of Man* [New York: Humanities Press, 1968], 140-144).

Even if a probabilistic origination of acts is no more impossible than the probabilistic landing of dice, a general aversion to mystery might motivate us to repudiate the whole notion of probability. If our aversion to mystery is not strong enough to motivate us also to repudiate self-determinism (with its infinite causal change and/or causing substances), we might then be moved to endorse moderate instead of extreme self-determinism. And, even if we accept the mystery of probabilities in the case of die-tosses, we might be less willing to do so where human choices are concerned, because *we* are humans, and have an extraordinarily intense desire to understand ourselves, and might therefore be moved to endorse moderate instead of extreme self-determinism. But these considerations are certainly less powerful motivations to endorse moderate self-determinism than the sense of the impossibility, as opposed to mere mysteriousness, of probabilistic origination, that I have just indicated and attempted to justify.

Extreme self-determinists can propose (as they in fact do) that someone's having certain desires of certain strengths is what makes it probable that he will originate certain behavioral intentions rather than others. It doesn't matter whether the emergence of these desires is itself determined or only probabilistic – whether his personality is a disposition (a determining tendency) to have certain desires or whether it is a propensity (a non-determining tendency) to have certain desires. This doesn't matter, because if the emergence of these desires is only probabilistic, this would be no more troublingly mysterious than the conformity of die-tosses to a probabilistic pattern. The mystery of probability becomes especially troubling only where agent-originated intentions are concerned, but at this point we aren't considering the probability of agent-originated intentions; we are only considering the probability that there will be certain desires of

certain strengths. The relative strengths of someone's desires would make it probable that he originate certain behavioral intentions rather than others either by disturbing a universal "probability field" that is external to him, or by conferring upon him a private propensity to originate certain behavioral intentions rather than others. But now *this* application of probability, at the level of the origination of intentions, *is* especially troubling, as the application of probability at the level of desire-formation wasn't.

Note that if extreme self-determinism were the position that our personalities play no role at all in explaining our behavior, then extreme self-determinists could not hold that we have propensities to originate certain intentions because our personalities give us these propensities. However, few if any self-determinists go so far as to deny that our personalities play any role at all in explaining our behavior, and so I give the name "extreme self-determinism" not to this *more* extreme, but seldom (if ever) maintained position, but rather to the more moderate position that our personalities can't prevent us from originating behaviors that aren't precluded by formal, environmental, and skill-related factors.

Self-determinists must not only explain the fact that groups defined by the natural backgrounds of their members display different behavioral profiles. They must also explain the fact that groups defined by the intentions that their members have originated also display different behavioral profiles. A group of people who have in the past originated intentions to live in an industrious rather than profligate manner will contain a larger fraction of members who will live in an industrious rather than profligate manner in the future than does a group of people who have in the past originated intentions to live in a profligate rather than industrious manner. This is a problem for extreme self-

determinists because it seems that if someone can originate an intention to live in an industrious rather than profligate manner at one moment, then he can originate an intention to live in a profligate rather than industrious manner a week later, if not a moment later. Yet we expect that someone whom we believe to have originated a certain kind of act in the past will still be acting in this way in the future. For example, we trust people who have kept promises in the past, and we don't trust people who have not kept promises in the past; that is, we expect that people who have kept promises in the past will still be promise-keepers in the future, and we expect that people who have not kept promises in the past will not keep promises in the future. If we believe that any intentions are ever originated (and most of us do believe this), then we must believe that at least sometimes people who choose to keep or break promises have originated these intentions, for these are morally significant intentions, and morally significant intentions are the ones we are most inclined to regard as agent-originated. We trust people who have usually kept their promises in the past, without worrying about whether or not they have originated their keeping of these promises. We wouldn't trust someone any less if we were to learn that he has originated his previous promise-keeping. So, we must think that the class of people who have originated their keeping of their promises contains more members who will continue to keep their promises than does the class of people who have chosen not to keep their promises. It is just as difficult for extreme self-determinists to explain why groups defined by the intentions that their members have originated display different behavioral profiles as it is for them to explain why groups defined by the natural backgrounds of their members display different behavioral profiles.

Self-determinists must explain both natural background-based group behavioral profiles and origination-based group behavioral profiles. Extreme self-determinism would have to explain both by attributing probabilities to originations of acts. If it is indeed impossible for originations of acts to have probabilities then extreme self-determinism is unacceptable. Unlike extreme self-determinists, however, moderate self-determinists *can* allow that our personalities could necessitate our behaving in certain ways – as long as there are limits to the extent to which our personalities necessitate this. So, unlike extreme self-determinists moderate self-determinists don't have to hold that it is probable that people with certain natural backgrounds will originate certain intentions. They can hold that as long as someone's origination of an act is a viable option, it is neither more nor less likely than not. The cracks between necessitations of behavior by personality are void of any degree of probability whatsoever.

Attacks upon self-determinism usually assume that self-determinism must be extreme self-determinism, and then urge that self-determinists cannot adequately explain group behavioral profiles.⁷¹ If self-determinism were indeed simply extreme self-determinism, then these attacks would succeed. But self-determinism can be moderate, and moderate self-determinists *can* explain group behavioral profiles. Therefore, these attacks fail. The moderate self-determinist solution to the explanation of group profiles is that a person's natural background causes him to have a certain kind of personality (a

⁷¹ For example, see Owen Flanagan, *The Problem of the Soul: Two Visions of Mind and How to Reconcile Them* (New York: Basic Books, 2002), 151-153. Flanagan also cites our presumption that we can influence people to become better or worse through persuasion and the example of our behavior, but this presumption only manifests our predictive confidence, which in turn rests upon our belief that people's behavior is generally determined by desires, and beliefs about how these desires may most effectively be satisfied, that are themselves determined by their heredity and past experiences. Our persuasive efforts, and the example that we provide becomes part of the past experience of the person we attempt to influence. We add these elements to the person's experience because we believe that when his experiential package is thus modified, his desires will be modified in a way that will make him behave as we desire him to behave.

disposition to have desires of certain kinds and certain intensities in certain situations), which in turn causes him to behave in a way that conforms to certain limits, given the state of his environment, his skills, *and his limited ability to originate a new desire or strengthen an existing one*. Moderate self-determinism can accept the mystery of psychological probability without accepting the apparent impossibility of a probabilistic origination of acts by allowing that someone's personality only probabilistically disposes him to have desires of certain kinds and certain strengths in certain situations. Someone would originate the placement of his behavior within the limits fixed by his desires. Here it may be noted that the placement of a moderately self-determining agent's acts within the limits established by his desires would be no more and no less predictable if he had no origination ability but were subject to a certain degree of random fluctuation of desire. Self-determinism and indeterminism would entail the same empirical consequences, although very different moral consequences.

I have introduced four forms of self-determinism: (1) extreme amoral self-determinism; (2) extreme moral self-determinism; (3) moderate amoral self-determinism; and (4) moderate moral self-determinism. The *initial*, "psychological", attractiveness of these forms of self-determinism is inversely proportional to their restriction of this range of acts that we can originate, with extreme amoral self-determinism being least restrictive and therefore most attractive, and moderate moral self-determinism being most restrictive and therefore least attractive. It isn't clear whether extreme moral self-determinism or moderate amoral self-determinism would be less restrictive and therefore more attractive. That would depend upon the frequency with which someone thinks that a possible act is

more right than the one toward which he is inclined. The more frequently this occurs, the less restrictive and more attractive extreme moral self-determinism would be.

I have already presented, at some length, an abstract, non-psychological consideration that nevertheless renders extreme *or* moderate moral self-determinism preferable to extreme *or* moderate amoral self-determinism: that it is only in the midst of moral deliberation, or while vividly recalling it, that we must believe that we can originate intentions to do things. It can now also be seen that *moderate* moral self-determinism is preferable to all of the other, less restrictive forms of self-determinism *because* it is more restrictive, even though this is what renders it less psychologically attractive. Moderate moral self-determinism's greater restrictiveness makes it more compatible with the usual correlation of personality with action, and with the behavioral predictability that depends upon this correlation. This gives us a strong reason to prefer it, even to extreme self-determinism. Unlike extreme self-determinism, moderate self-determinism is compatible with the predictability even of many morally significant decisions. I will now argue that the psychological needs that move people to embrace self-determinism, despite its paradoxes, even when they are not involved in moral deliberation, may be adequately met by moderate moral self-determinism.

Extreme amoral self-determinism is more attractive than the other, more restrictive, forms, first of all, simply because it depicts us as more powerful, and we prefer power to impotence. An originaive power is especially glorious, and so we especially prefer the unrestricted possession of such a power. Secondly, extreme amoral self-determinism is more attractive because it depicts us as deploying our originaive

ability in a greater number of actual circumstances, and there is something glorious about the deployment of originaive ability, beyond the mere possession of this ability.

I concede that this pair of considerations really does render extreme amoral self-determinism more attractive than the other forms, including the moderate moral self-determinism that I endorse, with respect to the quantity of self-admiration that it justifies. However, moderate moral self-determinism's restriction of our originaive ability and lessening of the number of actual circumstances in which we can deploy this ability is not as drastic as it might at first seem to be. This is because most of our activity is subject to determinations of degree; little of it is of the either-or type. We can be more or less diligent, more or less violent, more or less helpful. And, even where activity is of the either-or type, as is the case when we can either keep a promise or not, tell the truth or not, follow a rule or not, we can still originate choices regarding the extent to which we shall follow a general policy of keeping promises, telling the truth, and following rules.

Moreover, we can originate a more or less intense commitment to a certain general way of life, with its order of priorities, within the range of ways of life that are possible for us, and within the range of intensities of commitment to those ways of life that are possible for us. A moderate moral self-determinism must hold that we can always do this, for we can always consider the comparative rightness or wrongness of possible ways of life, and it seems that any way of life may be lived more or less intensely. Even if the strength of our desires and the weakness of our wills doesn't leave us much wiggle-room where living in a certain way more or less intensely is concerned, surely we can always originate a *little* more or less intensity. (We can originate a bit of de-intensification in our way of living by originating intentions to do things that don't

quite belong to this way of living.) Of course, we are not always thinking about how we shall live. But it is safe to say that we *must* frequently deliberate about whether to live as we are inclined to live or in another way that seems more right, and that when we are not thinking about how we shall live we are nevertheless living in a way that has been, or might have been, influenced by choices emerging from those past moral deliberations. A moderate moral self-determinism would hold that we either originate these choices or might have originated others. Since we must frequently engage in moral deliberation over how we shall live, we must frequently find ourselves in situations in which we originate intentions to live in a certain way (or to do so more or less intensely) or refrain from originating such intentions when we might have done so.

There are four more reasons for the greater attractiveness of extreme amoral self-determinism, as compared to the other forms of self-determinism, to which successful responses on behalf of the other forms, and especially on behalf of moderate moral self-determinism can be made. First, we think that the truth of extreme amoral self-determinism is required for the justification of our normal practices with regard to rewarding people for some of their acts and punishing them for others. Second, we think that our acceptance of any other position would paralyze us, because our ability to engage in normal, everyday deliberation requires that we endorse extreme amoral self-determinism. Third, we think that our being enduring, continuous “selves” requires its truth. Fourth, we think that our genuine involvement in the world requires its truth. I will argue that our endorsement of even moderate moral self-determinism is fully compatible with all of these things that we care about.

The first of these reasons for preferring extreme amoral self-determinism, that its truth is required for the justification of our normal punitive and reward-bestowing practices, depends upon the assumption that someone's moral responsibility for his act requires his having originated it or his having been able to originate another one. If someone is not morally responsible for his act then it is usually wrong to punish him for it, no matter how wrong the act is, where the "punishment" of someone involves the intention to make him suffer. So, it is concluded that if someone hasn't originated his act, and wasn't able to originate another one either, then it is wrong to punish him for it. The same concern applies to rewarding people for acts that they have not originated, although not as urgently; the problem is just that if people who don't deserve it get rewarded, then the rewarding of people who *do* deserve it will mean less. Now, if we hold that sometimes people's personalities make it necessary that they act in a certain way or prevent them from doing so, then we can never be sure that the people whom we want to reward or punish for their acts have originated these acts, or that they might have originated others. So, if we think that moral responsibility requires originaive ability, we will be inclined to endorse an extreme form of self-determinism. And, since many of the people whom we want to reward for their acts may not have performed these acts for the sake of their rightness, while many of the people whom we want to punish for their acts may not have thought that they should do anything else, we will be inclined to endorse extreme amoral self-determinism.

This reason for finding extreme amoral self-determinism most attractive has already been addressed. Someone's moral responsibility for his act doesn't require his having originated it or his having been able to originate another, better one. Only his

pure moral responsibility for the act would require this. All that is required for moral responsibility is the ability to form intentions expressive of one's personality and to form reasonable beliefs about the rightness or wrongness of possible acts. So, someone's not having originated his act, and not having been able to originate another one either, doesn't suffice to make it wrong to punish or reward him for it. This isn't enough to prove that his being morally responsible for it suffices to make it right to punish him or reward him for it. Something more needs to be said about that. But, given that moral responsibility for an act doesn't require the origination of it or the ability to originate another one, it isn't clear why the origination of an act or the ability to originate another one would make an agent more deserving of reward or punishment. Perhaps he would be more deserving of some divinely administered "eternal" punishment – but that is not *our* concern. Even the rejection of self-determinism altogether doesn't do anything to diminish the justice of our normal reward-bestowing and punitive practices, so moderate moral self-determinism is certainly not at a disadvantage here.

The second reason that we have for wanting extreme amoral self-determinism to be true is that we think that our rejection of it would entail our general paralysis. Normal life requires deliberation (usually not moral) over what to do, and someone's deliberating over what to do seems require that he think of himself as being able to originate any act that seems to him during the course of his deliberation to be a viable option for him. If we don't think that we can originate the acts over which we deliberate, then instead of deliberating we will just wait to see what we end up doing. But if we take this attitude, then we will never end up doing anything. (People commonly have this experience when they are in some comfortably passive state – lying in bed in the morning, or sitting on a

park bench – which they know they will soon have to bring to an end. They idly wonder, “Will I get up now? Okay, how about now?”) Most of our deliberations are nonmoral. So, it seems that acceptance of moral self-determinism would produce an almost constant paralysis, because instead of entering into these nonmoral deliberations we will just assume a paralyzing wait-and-see attitude. But even if we were to accept moderate amoral self-determinism we would be paralyzed. Moderate amoral self-determinism, like moderate moral self-determinism, entails that our personalities often make it impossible that we do or refrain from doing something. But whenever we aren’t sure that we can do something we can’t seriously deliberate over it. So, unless we are sure that an act is one of those that our personalities permit, we will not be able to deliberate over it. We can never be sure of this, so we will never be able to deliberate. We will always adopt a wait-and-see attitude instead, and as a result will be continually paralyzed.⁷²

This concern arises from the mistaken assumption that our endorsement of a metaphysical proposal will lead to our constant preoccupation with it. It is true, I think, that when the thought occurs to me that I can’t originate any of the acts over which I am deliberating I must then either stop thinking of myself as approaching a decision or must assume the attitude of wondering and waiting to see what I will decide to do. And, it is true that while I am wondering and waiting to see what I will decide to do I can’t decide to do anything. However, it doesn’t follow that I will never decide to do anything, for the thought that I can’t originate any of the acts over which I am deliberating will rarely occur to me. Our engagement in *moral* deliberation *does* seem to require our thinking (at least in some vague, back-of-the-mind sort of way) that we can originate an act that

⁷² Galen Strawson paints a persuasive picture of someone who attempts to integrate his rejection of self-determinism into his everyday life as paralyzing himself (*Freedom and Belief*, 102).

seems to us to be more right than the one that we are inclined to perform. But our engagement in non-moral deliberation, while *compatible* with the thought that we can originate any of the acts over which we are deliberating (and incompatible with its negation), doesn't require this thought. So, we can endorse moderate moral self-determinism without radically altering our daily lives.

Our third reason for wanting extreme amoral self-determinism to be true is that our "sense of self", or at any rate of an enduring, continuous self, seems to be dependent upon our belief that we can always originate our intentions. This concern may take two different forms. First, we want to maintain our ever-present, if usually only implicit, belief that we are simple substances, or at any rate don't want to accept anything that seems to contradict this belief. Apparently, only our ability to originate intentions can justify the thought that we are simple substances over and above (or under and beneath) the current, or swarm, or squabbling congress of our desires and thoughts (different metaphors seem appropriate at different times).⁷³ But according to anything other than extreme amoral self-determinism, we often can't originate our intentions; we must do what we are inclined to, or at most either what we are inclined to do or what is more right. So, if we accept a self-determinism that isn't extreme and amoral, we are pressed to conclude that we aren't simple substances between the times when we can originate our intentions. Somehow, during these intervals, only the current, or swarm, or congress, of desires and thoughts is left.

⁷³ Galen Strawson imagines a Buddhist monk who successfully integrates into his everyday life, through meditation, his theoretical repudiation of the idea of himself as a substantial self, as surrendering the idea of himself as a self-determining agent (Ibid., 119). What really threatens us, though, is the opposite possibility – that we might lose our sense of substantial self by meditatively integrating into our ordinary consciousness a theoretical repudiation of the idea that we are self-determining agents. Perhaps if we were to integrate into our ordinary consciousness a non-self-determinist understanding of decision of the sort proposed by George Ainslie we would have to take to heart his conception of "the self as a population" – a population of "contradictory preferences" – that is, as no self at all (*Breakdown of Will*, 39-40).

The conclusion entails various paradoxes: for example, that there might be intermittent phases of an identical substance. We need not address these paradoxes, however, for the conclusion really isn't entailed. For, first, our desires and thoughts are not entities apart from us, which might therefore be present even when we are gone. Rather, they are our own states, or events of which we are the subjects. But second, even if this weren't so, even if they *were* entities apart from us, then even according to moderate moral self-determinism we would nevertheless continuously retain the ability to originate intentions even when we can't deploy this ability. So, during the intervals between our originations of intentions we would just be sitting back, as it were, while the current of desires and thoughts flows past us, waiting for a time when it will seem to us that a possible act is more right than the one that we are inclined to perform. So, even if our being simple, substantial selves really does require our ability to originate intentions, acceptance of moderate moral self-determinism would be compatible with, and indeed would support, belief in a continuous self.

Our fourth reason for wanting extreme amoral self-determinism to be true is that it seems to us that our genuine, ongoing involvement in the world requires its truth. This worry assumes that we retain our substantial selfhood even when we aren't presently originating intentions, if only because we retain our undeployed ability to do so, but that during these intervals what we call our acts really aren't our own. They are produced by our desires, which stand, or sprawl, or flow apart from us. If it is even appropriate to call them acts, then our desires perform them; we don't. So, the involvement of the real self in the life that we ordinarily attribute to it would be intermittent at best if anything other than extreme amoral self-determinism were true, and most intermittent if moderate moral

self-determinism were true. When we aren't originating our acts, we are only sitting back and watching as our lives flow past us. Even an extreme amoral self-determinism according to which the origination of an intention requires the overcoming of inclination would be unsatisfactory, since in this case we would be uninvolved in the world whenever we do what we are inclined to do. Only our origination of *all* of our intentions – the doctrine of a doubly extreme amoral self-determinism – would be satisfactory.

My response to this concern, of course, is to again repudiate the notion that someone's desires are entities apart from him, whether inhabiting him, or pushing him around from the outside, or flowing past him. The thought that genuine involvement in the world requires an ability to originate acts is based upon this false assumption. If our desires are only our states, or are continuing events of which we are the subjects, then our involvement in the world does not depend upon our ability to originate our acts. Every time it is asserted that someone's desires cause him to do something, his involvement in the world is asserted, for it is through this very causing of acts by his desires (by his desiring) that he involves himself in the world. So, even the rejection of self-determination in general shouldn't threaten our sense of involvement. The acceptance of moderate moral self-determinism certainly shouldn't threaten our sense of involvement.

Moderate moral self-determinism can give us enough of what we want to be acceptable, then, even without our having any special reason to prefer it to extreme amoral self-determinism. But, as I have noted, we have two special reasons to prefer moderate moral self-determinism to extreme amoral self-determination. First, our experience of moral deliberation is all that ever requires us to believe that we originate our intentions. This renders moral self-determinism in general preferable to amoral self-

determinism. Second, moderate moral self-determinism's greater restrictiveness makes it more compatible than the other, less restrictive forms of self-determinism (including extreme moral self-determinism) with the probability that we will behave, and intend to behave, in certain ways rather than others.

8. *Previous self-determinist discussions of behavioral probability*

In this chapter, I will first review a few published extreme self-determinist discussions of behavioral probability, both amoral and moral to show that they either explicitly or implicitly attribute probabilities to our origination of intentions. Extreme *amoral* self-determinism makes all acts impervious to likelihood of any sort, while extreme moral self-determinism makes only morally significant acts impervious to likelihood. I will then review the only moderate self-determinist proposal regarding the problem of behavioral probability that I have encountered in formal philosophical literature – a proposal that happens to be an acknowledged misreading of one writer by another, and is endorsed by neither of them. This is a moderate amoral self-determinist proposal. I have discovered a moderate *moral* self-determinist approach to the problem proposal only in a popular book by C.S. Lewis.

According to Randolphe Clarke's extreme amoral self-determinism, we remain, despite our ability to originate our behavior, fully involved in the "causal order" of nature.⁷⁴ Clarke proposes that "an agent's behavior, besides being caused by her, is [probabilistically] caused also by earlier events, among which are her having or coming to have certain beliefs, desires, preferences, aims, values, and so forth."⁷⁵ Here "preferences and aims" are easily assimilated to "desires", and "values" might as well be, since they aren't decisively distinguished from non-evaluative motivating factors that operate as efficient rather than final causes. To value something, for Clarke, seems to be a sort of desiring of it that is distinguished from desirings that aren't valuing only by the special way the object is marked out for the desiring person: it has a special sort of value-

⁷⁴ "Toward a Credible Agent-Causal Account of Free Will", in *Nous* 27:2 (1993), 191-203; 194.

⁷⁵ *Ibid.*, 193.

coloring for him. Since it is a “probabilistic causing” that Clarke has in mind, Clarke would no doubt agree that the probability that we act in a certain way is proportional to the intensity of our desire to act in that way. Since Clarke holds that “an agent’s behavior” is probabilistically caused not just by her desires but *also* by her, Clarke seems to have in mind that the relative intensities of our desires make it probable that we behave in a certain way by making it probable that we will originate our behaving in a certain way. Since the relative intensities of our desires are what make it probable that we will originate our behaving in a certain way, our desires can be seen as “probabilistically causing” the behavior, even though we originate it. This, at any rate, is what Clarke seems to mean when he proposes that an agent-originated action is “probabilistically caused” by “earlier events”, so that “until [the agent’s] performance of that action, the chance that she would perform it remained somewhere between zero and one.”⁷⁶

Suppose, then, that there are two acts, A and B, that are possible for someone, and that his desires give a probability of .99 to his origination of A and a probability of .01 to his origination of B. He can still originate B then, and if he does so B will still also have been probabilistically caused by his desires.

At most, our desires and the apes’ desires would differ in their ultimate causal power only in that while apes’ desires deterministically cause their behavior, our desires do so only probabilistically. Clarke might allow, though, that apes’ desires, like ours, cause their behavior only probabilistically. The difference between us and the apes would only be that while the apes’ desires probabilistically cause their behavior *directly*, ours would do so by giving us propensities to originate the desired behavior.

⁷⁶ Ibid., 193.

Unlike Campbell's account, Clarke's account limits the range of behaviors that an agent can originate to those that he thinks will be at least slightly conducive to the satisfaction of at least one of his desires. An agent's supposition that an act will be conducive to the satisfaction of one or more of his desires is what confers some degree of probability greater than 0 upon his performance of it. (An agent's overall inclination to perform an act would presumably confer a probability of .99 upon it.) In the absence of any supposition that an act would be conducive to the satisfaction of some desire, then, the behavior cannot occur. Clarke's self-determinism is therefore not as complete as Campbell's, but it is still pretty extreme, since the range of our desires is vast. I can experience a slight desire to jump down onto the subway tracks and touch the third rail, for example, or to grab someone's baby.

Timothy O'Connor's extreme self-determinist proposal apparently agrees with Clarke's in holding that our prior desires confer upon us propensities to originate acts that we think will be conducive to the satisfaction of these desires. O'Connor seems to disagree with Clarke only with regard to whether or not these desires still operate as a redundant act-causing system, so that our desires continue to probabilistically cause our acts even as we originate them. Clarke endorses this redundancy. O'Connor rejects it.

It might seem inappropriate to attribute to O'Connor a crude psychology in which the factors giving us a propensity to behave as we do are all reduced to the common element of "desire." He speaks rather of a "nested structure of conative and cognitive factors," and of "the structure of my preferences, stable intentions, and so forth."⁷⁷ However, any "conative factor" is surely some sort of desire, and it seems likely that in mentioning "cognitive" along with "conative" factors O'Connor is only thinking of

⁷⁷ *Persons and Causes*, 95, 96.

beliefs that we have regarding how these desires might be satisfied. And, since O'Connor gives no hint that he sees a "preference" as possible in the absence of any "conative factor" it seems likely that in his view something would be preferred only if regarded as satisfying or as conducive to the satisfaction of some desire. As for someone's "stable intentions," it is hard to see how these could give him a propensity to do what he intends to do that is greater than his propensity to originate the doing of something else unless (1) his intentions match his inclinations, reflecting the overall configuration of his desires, or (2) he is inclined to do what he intends to do. (Even if this reductive psychology is not in fact attributable to O'Connor, what I have to say about him can still be read as aimed at a conveniently altered version of his proposal.)

"[I]t is a truism," O'Connor recognizes, "that given the structure of my preferences, stable intentions, and so forth, along with the situation with which I am faced, I am often far more likely to act in one particular way than in another."⁷⁸ Since O'Connor is an extreme self-determinist, this requires him to hold that the configuration of someone's desires confers upon him a greater propensity to originate certain acts than to originate others. On O'Connor's account, what enables someone to originate an act is that a desire gives him a *reason* to act. It is simplest to read O'Connor as holding that the agent's desires directly confer upon him greater or lesser propensities to originate acts for these reasons. O'Connor tells us that the agent's psyche is "a structured, dynamic situation open to some possibilities and not to others," and that the "the nested structure of conative and cognitive factors as 'my reasons' is among the most basic factors that circumscribe my capacity to exercise a limited degree of autonomy."⁷⁹ We can conclude,

⁷⁸ Ibid., 96.

⁷⁹ Ibid., 95.

then, that if someone doesn't think that doing something will be even slightly conducive to the satisfaction of any of his desires, he won't have a reason to do it and therefore won't be able to originate doing it. So, O'Connor's self-determinism, like Clarke's, is somewhat restricted by the requirement that someone must think that an act would be at least slightly conducive to the satisfaction of a desire of his if he is to originate it. Again, though, the restriction would still be so minimal, given the vast range of desires that we can feel to some slight extent, that it should be called an extreme self-determinism.

We originate an act, according to O'Connor, by "recognizing a reason to act" and in recognizing a reason to act we originate an intention to act in that way. The origination of this intention to act was something that we had a certain propensity to do: "[T]he agent may be said to carry a propensity to *generate* a specific type of intention. We may suppose, that is, that recognizing a reason to act induces or elevates an objective propensity of the agent to initiate the behavior."⁸⁰ Surely, O'Connor doesn't mean to suggest that our reasons form a distinct layer of psychic entities intervening between our desires and propensities. It isn't as though our desires generated reasons which we would perceive before "recognizing" them, and the comparative "weights" of these reasons would then give us our various propensities to recognize them. Rather, our desires, and our beliefs regarding what would lead to their satisfaction, *are* our reasons to act, or become so when we recognize them as such.

O'Connor clearly holds, with Clarke, that we have a natural system of desires (or "conative and cognitive factors", at any rate), upon which our originative ability is superimposed. He just as clearly holds that this natural system of desires confers upon us our propensities to originate certain kinds of action. However, he rejects Clarke's

⁸⁰ Ibid., 97.

“attempt to have an irreducible agent causation fitted into or on top of an unbroken chain of event causation.”⁸¹ (O’Connor assumes that the origination of acts must consist in agent causation, and that event causation cannot be originaive.) Does O’Connor hold, then, that someone’s natural system of desires confers no direct behavioral propensities upon him? Or, does he hold that someone’s natural system of desires does confer behavioral propensities upon him, but that as soon as he exercises his ability to originate his acts these prior behavioral propensities are neutralized? It is easier to understand him in the second way, which is in any case more plausible because it better preserves our kinship with our ape ancestors. Thus agent causation is not fitted “on top of an unbroken chain of event causation” (the causation of acts by the sorts of propensities to act in certain ways that we would attribute to nonhuman mammals) in O’Connor’s account because in this account the chain of event causation is *broken* as soon as we originate an act. We simply overrule the chain of event causation.

Clarke and O’Connor don’t indicate that they conceive of the origination of an intention to do something as in any way altering the balance of someone’s desires so as to make it more likely that he will choose to behave in this way in the future. O’Connor’s account in particular doesn’t invite such a proposal, since it pictures the origination of an intention as transpiring through the “recognition of a reason to act,” where this reason to act would be provided by a prior desire. In recognizing a reason to act, someone is merely aligning himself with the corresponding desire. The proposal that this alignment of oneself with a desire would strengthen this desire seems arbitrary.

⁸¹ Ibid., 79.

Clarke and O'Connor simply accept the notion that it can be more or less probable that we originate an act, or an intention to behave in a certain way. That the origination of an act or intention should have any degree of probability at all seems impossible to me.

Robert Kane and Laura Waddell Ekstrom have separately offered what they regard as self-determinist models accounting for the probability that people will do what they most intensely want to do. Kane and Ekstrom explicitly reject the agent-causal account of the origination of acts according to which an agent's origination of his act consists in his being the first cause of it. They identify an agent with a *somewhat randomly* evolving system of desires and propensities to have certain desires. Here I am once again, for the sake of convenience, reducing the various constituents of a subtle psychology to species of "desire". But I don't believe that in doing so I deform the Kane-Ekstrom proposal to any great extent. Ekstrom's version of the proposal centers upon the image of a "character system," in which the distinction between a present desire and a propensity to have such a desire is glossed over, just as it is in our ordinary thinking about desire, when we say that we have "always wanted" this or that:

Suppose we take an agent to be constituted by a character, together with the power to fashion and refashion that character. A character, or *character system*, is an aggregate of preferences and acceptances. Then an agent is an evaluating and choosing faculty (by which she creates preferences and acceptances), along with the character system, made up of those preferences and acceptances.⁸²

The key point here is that "[t]he decision concerning what preference to form ... is caused, but not fully determined, by the past."⁸³ So the "choosing faculty" that combines with a "character system" to constitute an agent turns out to be nothing more than a somewhat randomly evolving process of preference-(and acceptance)-formation. Given

⁸² Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Westview Press: Boulder, 2000), 113.

⁸³ *Ibid.*, 115.

that preferences are surely what we have when we prefer things, and as such involve a disposition to do what is preferred, and that acceptances are similarly what we have when we accept things, and as such involve a disposition to put up with, or continue doing, what we accept, it seems perfectly acceptable, if not preferable, to regard Ekstrom's preferences and acceptances as species of desire.

Kane speaks of a "self-network" rather than of a "character system," and (like O'Connor) of "reasons" rather than of "preferences and acceptances":

When we say that conflicting reasons sets R and R' motivating [someone] are both *hers* – parts of her general motivational system – this corresponds, in neural terms, to saying that the neural connections representing R and R' are imbedded in a more comprehensive network of neural connections representing the general motivational system in terms of which she defines herself as agent and practical reasoner. I propose to call this comprehensive network representing her general motivational system the "self-network."⁸⁴

Since Kane's "reasons sets" *motivate* people, it seems acceptable to identify them with combinations of desire and beliefs regarding what might be conducive to the satisfaction of these desires. Just as in Ekstrom's version of the proposal the agent's self-determination consists in the degree of randomness with which his character system evolves, so it is in Kane's version:

[C]onflicts of will... stir up chaos in the brain and make the agents' thought processes more sensitive to undetermined influences. The result is that, in soul-searching moments of moral and prudential struggle, when agents are torn between conflicting visions of what they should become..., the outcomes are influenced by, but not determined by, past motives and character.⁸⁵

In both versions of the proposal, someone's undetermined decision is a spontaneous alteration of the "character system" or "self-network" that constitutes him. This

⁸⁴ Robert Kane, *The Significance of Free Will* (Oxford University Press: New York, 1998), 139.

⁸⁵ *Ibid.*, 130.

alteration of the character system or self-network had a certain degree of likelihood, generated by the previous unstable configuration of someone's desires, together with his beliefs regarding how these desires might be satisfied.

Since the identification of an agent with *his* own mental states and propensities is incoherent, it is reasonable to suppose that Kane and Ekstrom would regard these propensities as inhering in a human body. (An immaterial substance, or soul, would do just as well, but Kane's interest in the indeterminacy of subatomic events within the brain strongly supports a reading of him as thinking of "our" desires and conative propensities as states and propensities of our bodies. Ekstrom, also, exhibits an anti-metaphysical bias.) The agent is then the system of "his" body's desires and conative propensities. (We are assuming that a body can be a subject of mental states.) He – this system – evolves *randomly*, to a certain extent, although for the most part under the guiding influence of the body's heredity and its interactions with its environment. According to the Kane-Ekstrom account, it is never entirely *impossible* that even the weakest propensities of a personality will prevail, so that he begins very often to very much want what he has never very much or very frequently wanted before. The character-system or self-network (which I will call the "self-system", so as to combine the two labels) might suddenly evolve in a way that involves the sudden strengthening of previously obscure propensities. This would account for conversion-experiences.

Since the prior, indeterministically-established configuration of the self-system cooperates with the body's heredity and its interactions with its environment to establish the probabilities of its evolving in various ways, it can be said that a self-system contributes to the indeterministic causing of its evolution in a certain direction as new

intentions form within it. The Kane-Ekstrom account can therefore be regarded as a substance-causation account of the origination of acts. The agent is not the sole first cause of his decisions, since the body's heredity and environmental interactions play a large role in causing them as well, but he is *a* first cause, and that is what matters. It may be objected that the whole process of indeterministic evolution is kicked off by the body's heredity and initial interactions with its environment, and that the self-system is therefore not a *first* cause of its evolving as it does. However, this objection fails as long as this initial generation of a self-system by the body's heredity and initial interactions with its environment is itself indeterministic. If the self-system's further evolution is indeterministic, then its first emergence would be indeterministic as well, it seems. The self-system is in part the way it is to begin with for no reason at all. It just is what it is (to some extent), and then immediately begins to participate in the indeterministic causing of its own evolution.

Of course, the Kane-Ekstrom self-system is not a simple substance, as the self is under "traditional" agent-causal accounts of self-determination. But Kane and Ekstrom regard this as a virtue of their account: it doesn't depend upon the notion of an inner, spiritual being, or soul, unaffected by the laws of nature (including natural randomness), as constituting a person's core identity. Moreover, their representation of an agent's origination of his acts makes the generation of group behavioral profiles by propensities no more troubling than the generation of a certain overall pattern of dice-tosses by repeated random tosses of a die. Finally, the Kane-Ekstrom proposal has the advantage of explaining how the past origination of intentions can render the future onset of similar intentions more likely. The origination of a new sort of intention is the result of a

mutation of the self-system, and this mutation remains in effect, so that similar intentions will be generated in similar situations in the future, until further mutations of the self-system occur.

We might object to the Kane-Ekstrom account as failing to represent our self-experience – our sense of ourselves as being the single subjects of numerous desires and other mental states and acts, rather than as systems including these states and acts as constituents. However, I won't proceed in this direction. Rather, I repudiate the Kane-Ekstrom account of the origination of acts because it portrays the origination of an act as something that someone doesn't even *do*, let alone do intentionally. It is something that happens *to* someone, or within him. The Kane-Ekstrom account therefore provides no basis for attributing a purer moral responsibility to someone for an act that he has originated. Our wish to attribute a purer sort of moral responsibility to one another for some of our acts is our only motive for persisting in our initial belief that we originate them. The Kane-Ekstrom account can certainly explain group behavioral profiles, but only because it is not an acceptable self-determinist account.

At the end of Chapter 4, I summarized what I take to be C. A. Campbell picture of the originative process, a picture much like my own except that Campbell's agent plays the role of first cause, while mine is the subject of an beginningless series of originative acts. Campbell's "act of decision" takes the place of this infinite series, his "effort" takes the place of my inception or strengthening of a desire to do whatever is right, and his "will" is my "intention. Campbell's "effort of will" is the causing of intention by means of the infusion of a desire-like "energy" into the psychic economy, so that the old inclination is replaced by a new one, an inclination to do the right thing (although

Campbell does not call it that), which is then best understood as causing the intention to do the right thing. If an intention to do the right thing is not originated in this way, we ... allow our desiring nature to take its course.”⁸⁶ That is, we follow our old inclination, which presumably causes us to intend to do so. Campbell never speaks of a “will” to follow inclination, but in calling the will to do the right thing “effortful” he seems to distinguish it from a will that is *not* effortful, which would have to be the will to do what one is “naturally” inclined to do.⁸⁷

Campbell is an extreme moral self-determinist: he holds that we can *always* originate doing the right thing, by originating the “will” to do it. Lack of skill and environmental impediments wouldn’t be a problem, presumably, because the right thing to do could only be something that our skills and environment permit. He doesn’t envision any restriction by our personalities of our ability to do the right thing. Nevertheless, the requirement that our originated acts be doings of the right thing greatly restricts the fraction of acts for which we would have pure moral responsibility. Someone would be able to originate an act only when he is inclined to do something other than what he thinks more right. We usually don’t think that something is more right than what we are inclined to do. We usually don’t even think about the rightness or wrongness of what we are inclined to do.

The intensity of Campbell’s “efforts of will” (caused by agent-originated “acts of decision” and in turn causing the doing of the right thing) must be proportional to the strength of the desires that are thereby overcome: “We are all aware that a greater effort of will is required in proportion to the degree in which we have to transcend our formed

⁸⁶ *In Defence of Free Will*, 42.

⁸⁷ *Ibid.*, 49.

character in order to will the right.”⁸⁸ He describes this effort of will as the introduction of “a new energy.”⁸⁹ Apparently, then, he has in mind the deployment of some sort of force that is commensurable with that of the desires that it overcomes. This looks very much like the picture that I advocate, of the origination of or strengthening of a desire to do whatever is right.

In my summary of Campbell’s view at the end of the fourth chapter, I examined Campbell’s reasons for denying that his “new energy” is desire. Whether or not it merits the name “desire”, it is certainly desire-like. Campbell thinks that we can originate an unlimited quantity of this desire-like energy. We can make as great an “effort of will” as we need to make, given the strengths of our opposing desires. “[I]t rests with our self absolutely to decide whether we exert the effort of will which will enable us to rise to duty...” or whether we follow our inclination.⁹⁰

Campbell points out that that the moral character of his self-determinism makes most behavioral probability unproblematic.⁹¹ It is only when a possible act seems more right to us than what we are inclined to do that we are able to originate it. However, it is precisely in people’s making of morally significant choices that the correlation between the kinds of natural background that they have and the kinds of behavior that they choose are most glaringly evident. Appeals to the determining power of natural backgrounds are loudest in courtrooms, where they are made on behalf of people who find themselves there because they made morally significant choices. The defendant’s lawyer urges that

⁸⁸ Ibid., 47.

⁸⁹ Ibid., 42.

⁹⁰ Ibid.

⁹¹ Ibid., 46-47.

his client's childhood abuse and impoverishment made him such that he had to act as he did.

Campbell deals with this remaining core of problematic cases, in which someone's predetermined inclination makes it likely that he will follow it rather than doing the right thing, in a cursory, unsatisfactory manner, telling us that when we must make morally significant decisions,

...our character as so far formed prescribes a course opposed to duty, and an effort of will is required if we are to deviate from that course. But of course we are all aware that a greater effort of will is required in proportion to the degree in which we have to transcend our formed character in order to will the right. Such action is, as we say, 'harder'. But if action is 'harder' in proportion as it involves deviation from formed character, it seems reasonable to suppose that, on the whole, action will be of rarer occurrence in that same proportion: though perhaps we may not say that at any level of deviation it becomes flatly impossible.⁹²

I take that Campbell's "character" is the same as what I call "personality": someone's propensity to have certain desires of certain strengths in certain situations. Campbell's proposal that an "action is 'harder' in proportion as it involves deviation from formed character," follows from the thought that the more I would rather do something other than A, the harder it is for me to do A. The "steeper" someone's inclination – the more the desires that tend to lead him in that direction preponderate over those that tend to lead him in the opposite direction – the less likely it is that he will oppose it.

The description of an act as "hard" reflects a *feeling* of effort that we have in performing it. This feeling arises, I think, from our not wanting to perform the act itself. We may be inclined to perform it, because we want things that we hope to get by means of it, without wanting to perform it. But then we have this feeling of effort, of straining

⁹² *In Defense of Free Will*, 47.

forward, *whenever* we do something that we don't want to do. We don't have to be doing the right thing in order to feel this. We might have a sense of effort in studying the instruction manual for some piece of entertainment equipment, when the motivating desire (for fun) is thoroughly nonmoral. This is the same sense of effort that we might have in keeping a promise. Just as we would rather not read the instruction manual, so we would rather not keep the promise. But we are inclined to do both, because of the strength of our respective desires for fun and doing whatever is right.

It indeed seems that the more difficult we expect an act to be, the less likely we are to perform it. However, the case of the instruction manual shows us why expected difficulty and unlikelihood of performance are linked. The more difficult we expect an act to be, the more likely it is that the strength of our desire not to perform it will incline us against performing it. But this assumes that our origination ability may be restricted by the strength of our opposing desires – or rather, by the shape of our desire-generating personality. Campbell doesn't want to accept this assumption.

Campbell doesn't explain why our greater inclination not to do the right thing would make it less likely that we do it. He merely notes that “it seems reasonable to suppose” that there is such a relationship. A desire to avoid the pain of effort cannot be the reason; this would only be another “moral temptation”, to be overcome through the same moral effort produced by the same decision. So too, when we speak of “difficult” choices, we refer to the deliberative stress that precedes the effortless choice. It hurts to hover over the options; the descent is painless. The difficulty of our deliberations therefore cannot influence the frequency with which we originate acts for the sake of the goodness that we attribute to them.

If Campbell were to recognize the possibility of a standing desire to do whatever is right, and were to understand the origination of an intention to do the right thing as involving the origination or strengthening of such a desire, then he would be able to explain how the *past* origination of an intention to do the right thing can make it more likely that someone will choose to do to the right thing now and in the future. As we have seen, Campbell does not appear to recognize the possibility of a natural desire, whether momentary or enduring, to do whatever is right. He also doesn't appear to recognize the possibility that the new desire-like "energy" that the act-originating agent introduces into his psychic economy might linger in the form of an enduring (if gradually dwindling) desire to do the right thing. Since he doesn't recognize the possibility of such a desire, he can't explain why it is more likely that someone who has in the past originated an intention to do the right thing will now or in the future make a similar choice.

Thomas Reid champions what I have called a "bivalent" extreme moral self-determinism, according to which we originate the following of inclination as well as the doing of the right thing, or at least an intention to do either of these things. Reid envisions a "grand and ... important competition of contrary motives... between the animal, on the one hand, and the rational on the other."⁹³ Someone would originate the "prevalence" either of the strongest among the animal motives or of the strongest among the rational motives, making these motives effective.⁹⁴ The becoming-effective of his strongest animal motive or of his strongest rational motive would presumably be the

⁹³ *Essays on the Active Powers of the Human Mind*, 291.

⁹⁴ *Ibid.*

“determination of the will” that Reid envisions him as originating.⁹⁵ Whether this motive now *is* his “will” (or intention), or its prevalence causes him to will (or intend) what he does isn’t clear; my guess is that it is the latter. So, someone would originate either an intention to follow inclination or an intention to do the right thing.

Animal and rational motives are incommensurable; it is the agent who makes one or the other prevail. These “motives” look a lot like what we ordinarily call “desires”. So, Reid pictures an agent as causing either his animal desire or his rational desire to cause his behavior.

Although Reid sees his proposal as a defense of “liberty”, I would reframe it as one regarding what is required for pure moral responsibility. Reid’s proposal makes us always purely morally responsible for our acts. Since according to Reid we always have both animal and rational desires, and these animal and rational desires are incommensurable, it follows that only our determining of our will could enable us to act. Indeed, Reid embraces this consequence, advising that “[t]o say that man is a free agent, is no more than to say that in some instances he is truly an agent...” while “to say that he acts from necessity, is to say that he does not act at all.”⁹⁶ We can reframe this as a statement about pure moral responsibility rather than about freedom (or liberty). It then becomes the statement that to say that someone is purely morally responsible for his acts is no more than to say that he has acted, while to say that he has is not purely morally responsible for his act is to say that he has not acted at all.

Reid’s bivalent extreme moral self-determinism, according to which we originate intentions *either* to follow inclination *or* to do the right thing, doesn’t fit well with our

⁹⁵ Ibid., 265.

⁹⁶ Ibid., 280.

experience of resisting natural inclination in making ourselves finish a task, tell the truth, or apologize to someone that we have mistreated. We feel that if we were to let things slide – surely not the positive origination of what Reid might call an “animal” intention – we would let the task go undone, the truth go untold, the apology go unsaid. We wouldn’t then be *originating* this following of inclination.

Reid acknowledges that behavioral probability has been thought to be incompatible with the extreme moral self-determinism that he endorses. His brief, dismissive response to the problem of behavioral probability is to observe that “it may surely be expected” that foolish people will originate acts bringing them short-term gratification, while wise people will originate the acts that seem truly right to them.⁹⁷ This response is unsatisfactory. On Reid’s own account someone originates an act by originating his alignment either with his animal or with his rational motives. Someone’s alignment with his animal motives would constitute him as foolish; his alignment with his rational motives would constitute him as wise. It would therefore be someone’s origination of his act (by originating his own alignment with his animal or rational motives) that constitutes his foolishness or wisdom in the first place. There is no reason why someone who has previously been aligning himself with his animal motives shouldn’t suddenly begin to align himself with his rational motives, or, indeed, why he shouldn’t swing erratically back and forth between these contrary motives. He originates either alignment. So, there is no reason why, on Reid’s account, someone shouldn’t swing wildly and unpredictably back and forth between wisdom and foolishness. Nor would someone’s present wisdom or foolishness enable us to predict even what he will decide to do in the immediate future, since to attribute a present wisdom or foolishness to

⁹⁷ *Essays on the Active Powers of the Human Mind*, 292.

him is really only to characterize his most recent morally significant decision as having been made in favor of present gratification or in favor of what seemed right.

The only published moderate self-determinist proposal that I have found is sketched by Timothy O'Connor as a possible misreading of Randolph Clarke's proposal. This is an amoral self-determinist model: it does not present an agent's origination of an act as requiring his supposition that the act is more right than what he is inclined to do. O'Connor first summarizes Clarke's proposal:

[t]he agent's causal capacity consists in the ability to make effective an indeterministic propensity of one's reason to bring about A ... in the sense of 'acting alongside' or bolstering the tendency (whether it be of a low- or high-probability measure), ensuring that it will achieve its characteristic effect.⁹⁸

O'Connor then observes that "at first sight" this proposal

treats the agent's exertion of active power as a kind of additional force at work in the situation, one that has a measurable strength that might vary from situation to situation or person to person. ... Presumably these exertions would (even together with the reason-based tendency) sometimes be ineffective....⁹⁹

Clarke's own proposal thus seems "at first sight" to O'Connor to be the proposal that agent-originated choice always consists in an "exertion of active power" that can "sometimes be ineffective." The ineffectiveness of an exertion of active power would presumably be due to there not being enough active power exerted to get the job done. If not enough active power is exerted to get the job done, then presumably there wasn't enough active power available for exertion. "Active power" then seems to be what we ordinarily call "willpower". Again, willpower need not be envisioned as a sort of fuel-

⁹⁸ *Persons and Causes* 78, referring to Randolph Clarke, "Agent Causation and Event Causation in the Production of Free Action," *Philosophical Topics* 24 (fall 1996), 19-48.

⁹⁹ *Ibid.*

supply for a desire-machine that isn't running at full capacity. It is just an agent's ability to do something, or to intend to do something, that he isn't inclined to do.

O'Connor's objection to this moderate self-determinist model is that the notion of an ineffective exertion of active power is "intuitively inconsistent with the idea of agent causation". O'Connor asks,

How could it be that I exerted active power in forming my own intention, and the intention aimed at, so to speak, failed to have been formed? On what basis could it be said, even by myself, that any instance did involve a fruitless exertion – because I intended to form a certain intention?¹⁰⁰

My answer to O'Connor's question, "How could it be...?" is that the quantity of active power that I deploy would be insufficient to form an intention, to use O'Connor's terms. As for O'Connor's question, "On what basis could it be said...?", here he seems to be asking for a formal criterion of "fruitlessness" in the exertion of active power. The only formal criterion of fruitlessness in the exertion of active power that O'Connor can think of, it seems, is that someone would have an intention to exert active power, and would be unable to put this intention into effect.

O'Connor dismisses this proposal because it places intention before the "exertion of active power", rather than after. Indeed, O'Connor cannot accept the notion of an intentional deployment of willpower, because he is unwilling to accept an infinite regression of origivative events. He is unwilling to accept an infinite regression of origivative events because he endorses a substance-causal account of self-determination, according to which the agent must be the first cause of an act that he originates. If the agent must be the first cause of an act that he originates, then an infinite regression of origivative events would either (1) require the mysterious Chisholmian "nesting" of this

¹⁰⁰ Ibid.

infinite regression somehow *within* the agent's causing of his ultimate act, or (2) would be simply incoherent, since the agent would never find his place at the beginning of the chain of originative events, so as to render them originative. Because I endorse an alternate, event-causal account of origination that accepts an infinite regression of originations of intentions to originate intentions in which every originative event is an intentional act with the agent as subject, I do just what O'Connor objects to: I place intention before the exertion of an active power. However, I also place an exertion of active power before intention, and so on to infinity.

Still, I agree with O'Connor's rejection of the image of someone as failing to exert the active power that he intended to exert. I understand the fruitlessness of a fruitless exertion of active power as consisting not in the agent's failure to exert the active power that he had intended to exert, but rather in the insufficiency of the desire that this exertion of active power causes. The desire is insufficient if it cannot so alter the configuration of the agent's desires that he will now do the right thing that he was formerly disinclined to do. An either-or case (either the right thing is done or inclination is followed) would be rare, though. Far more common would be cases in which the origination of a desire, or of its strengthening, results in an incomplete improvement of the agent's behavior, at least in his own eyes.

The moderate self-determinism that emerges from O'Connor's acknowledged misreading of Clarke is amoral. As such, it is less acceptable than a moderate moral self-determinism, as any amoral self-determinism is less acceptable than a moral self-determinism. Because someone's origination of an act or intention would be utterly mysterious, we should be as restrictive as we can be in attributing the origination of acts

to people. Therefore, only great psychological and moral pressure can justify our endorsement of the proposal that someone can originate an act. We feel this pressure only when we are involved in, or vividly recall, moral deliberation. What we need, then, is a self-determinism that combines the moral orientation of Reid's and Campbell's accounts with the moderation of the one that O'Connor briefly sketches and rejects.

C.S. Lewis hints at something like a moderate moral self-determinism when he suggests that God may take our natural temperaments into account in judging us.¹⁰¹ The naturally ill-tempered person who makes a great effort to conduct himself in a minimally civil manner gets more heavenly credit than the natural sweet-tempered person who behaves exactly as his temperament disposes him to behave. It might be determined that a child will acquire a disposition toward violent behavior as he grows up, but the child may still make himself more or less violent. So, God credits or blames him only for the extent to which he approaches the upper or lower limits of his predetermined range of violence. However, proposals based upon this insight are inadequate so long as they do not include a plausible account of how the limits of the range of behavioral dispositions within which a person may situate himself are determined, and how it is that he may originate his placement within this dispositional range.

Although I haven't been able to find previous presentations of moderate moral self-determinism, other than this hint by C.S. Lewis, I am sure that, given its commonsensical character, hundreds of people have thought of it.

¹⁰¹ *Mere Christianity* (New York: Simon & Schuster Inc., 1996), 86-87.

9. *Summary of the proposal, and discussion of relevant issues*

Here is a point-by-point outline of the moderate moral self-determinist account of pure moral responsibility that I endorse. This account assumes that someone is able to have reasonable beliefs about the rightness and wrongness of acts that are possible for him. This is the second of the two personal requirements for any sort of moral responsibility, whether ordinary or pure. The second of the following points will show how the first of these requirements – that someone’s intention express his personality – is met. When I say that we “make ourselves” do, intend to do, or want to do something, I mean that we originate our doing, intending to do, or wanting to do it.

Point 1. Our pure moral responsibility for our intentions and acts may be positive or negative. We can have positive or negative pure moral responsibility for our intentions even when circumstances don’t permit our behaving as we intend to behave. We have *positive* pure moral responsibility only for those intentions and acts that we originate. We originate an act by originating an intention to behave in a certain way, and we originate an intention to behave in a certain way by making ourselves want something so forcefully (either originating a new desire or strengthening an old one) that the resulting new configuration of our desires causes us to intend to behave in that way. We have positive pure moral responsibility for intending to behave in a certain way if we have originated our intention to behave in this way, regardless of whether or not it turns out that we actually behave in this way. We have *negative* pure moral responsibility for acts of ours that we don’t make ourselves perform if we might have made ourselves act in another way. We have negative pure moral responsibility for our intention to behave in a certain way even when we haven’t made ourselves intend to behave in this way,

regardless of whether or not it turns out that we actually behave in this way, if we were able to make ourselves intend to behave in another way.

Point 2. When we originate a desire, or strengthen an existing one, in originating an intention to behave in a certain way, we modify our personality so we now tend to have such desires, or to have them at this new level of force. Our originated intentions therefore express our modified personalities, and so the first personal requirement for moral responsibility is met in cases of positive pure moral responsibility – that if someone is to be morally responsible for his intention, it must express his personality. If someone doesn't originate his intention, but might have originated another one, he meets the same personal requirement for moral responsibility as long as his unoriginated intention expresses his personality. So, this requirement may also be met in cases of negative pure moral responsibility.

Point 3. According to my event-causal account of origination, an originated act is the final term of a beginningless series of intentional acts of forming an intention to form an intention to, ultimately, behave in the way in question. I call these intention-forming acts "originative" acts. Trying to do something is an act. So, we might perhaps have positive pure moral responsibility just for *trying* unsuccessfully to make ourselves intend to do something. We also have negative pure moral responsibility just for not trying to make ourselves intend to do something, when we might have tried, even if the effort couldn't have been successful. Further, we might have negative pure moral responsibility for not trying as *hard* as we might have to make ourselves intend to do something

Point 4. When we make ourselves intend to behave in a certain way, we make ourselves intend to do the right thing: that is, we make ourselves intend to behave in this

way in order to do what seems more right than what we would otherwise have been inclined to do. Here the “in order to do” clause belongs to the description of what is intended. Thus, what we make ourselves intend is this: the doing of what seems more right than what we would otherwise have been inclined to do, *for the sake of* its greater rightness. But the origination of this intention is itself an act, and is therefore the result of an intention. So, when someone originates an intention to, say, keep a promise, his intention in originating this intention to keep a promise is, again, to do the right thing: keeping the promise for the sake of this deed’s rightness. The beginning of an intention, when other intentions were thought possible, is a decision. So, someone decides to make himself do the right thing, and at the same time he makes himself decide this.

Point 5. We deserve pure credit for originating an intention to do what we *reasonably* believe to be more right than what we would otherwise have been inclined to do, for the sake of its greater rightness. We deserve pure blame for not originating such an intention when we might have.

Point 6. The strengths of someone’s existing desires, and the quantity of his willpower, or ability to originate an intention to do something that he isn’t inclined to do, determine whether or not it is possible for him to originate such an intention. It is simplest to think of someone’s desires as having absolute, rather than comparative, strengths. A desire’s absolute strength might be measured by seeing how much effort someone would make to satisfy the desire, if it were the only one he had. This would require that effort be the overcoming of a tendency to remain at rest, or return to rest, which is not merely the effect of a desire to rest. When we say that we want to rest, we would have to be understood as expressing this tendency, rather than as expressing a

genuine desire. This tendency toward rest, which might be called “psychic inertia”, would have to be the same in everyone: it would be basic. Someone who “wants” to rest is then someone whose desires are insufficient to enable him to overcome this psychic inertia. If we assign this basic psychic inertia a value of 1, then we might assign a value of 1.1 to a desire that barely motivates someone to try to satisfy it, so that he makes a minimal effort to do so. Suppose, then, that this is someone’s only desire. If is able to originate an intention to do something else for the sake of its greater rightness, the quantity of his willpower is greater than 1.1. Typically, the satisfaction of some of someone’s desires would be incompatible with his doing what is right, the satisfaction of others would be compatible with his doing what is right, and the satisfaction of still others would require his doing what is right. Those desires whose satisfaction would require his doing what is right might include, but would not be limited to, desires to do whatever is right and whatever is best. Someone need not deploy all of his willpower in originating an intention to do the right thing. He need only deploy enough of it to modify the configuration of his desires enough to make them cause him to intend this. He can modify the configuration of his desires by making himself begin to desire, with a certain degree of force, to do whatever is right, or by making himself desire this more forcefully.

Point 7. In originating an intention to behave in a certain way, we conclude either a two-option or a multi-option moral deliberation. Examples of cases in which we are involved in two-option moral deliberation include those in which we have to decide whether or not to take a bribe, whether to go to graduate school or accept leadership of the crime family (where communal pressures make these the only two live options), whether or not to put the dollar in one’s pocket into a beggar’s cup (when one has no

coins, and the donation of a five dollar bill just doesn't come to mind as an option), or whether or not to attend the evening prayer service. In all of these cases, we are inclined toward one of the options, while the other seems more right – otherwise, the deliberation is not moral. An example of a multi-option moral deliberation would be one in which someone can follow his general inclination to a greater or less extent. Suppose Frank's general inclination is to assault Ted. Beating Ted would suit Frank most (he would follow his inclination all the way), but shoving Ted would suit him to a lesser extent, while a mere denunciation would suit him least. Nevertheless, Frank sees the denunciation as most right and the beating as less right.

Cases of two-option moral deliberation often depend, for their binary character, on the rather arbitrary exclusion of whole ranges of alternatives as being outside of the realm of normal consideration, or on a debatable definition of what is at stake. Why should it be outside the range of normal consideration that one might give a five or twenty to a beggar, or even give him one's coat or, for that matter, invite him home for a shower, pizza, and lots of beer? Only the arbitrary exclusion of alternatives prevents the question of what to do about the beggar from turning into a case of multi-option deciding. Moreover, two-option moral deliberations often require multi-option moral deliberations after a decision is made. Thus, after someone decides to attend the evening prayer service, and as result finds himself there, he must decide how much attention to pay to the prayers he recites, where his inclination is to race through them while careful attention seems most right to him. Two-option moral deliberations seem to play a more prominent role in theoretical discussions, probably because they are easier to deal with, but multi-option moral deliberations figure much more prominently in real, everyday life. Usually,

when we have to make a morally significant decision, we have to choose from among more than two options. Here, we decide upon some point along a continuum of practical intensity where greater practical intensity is perceived as more or less right.

These would be someone's *behavioral* options. Even in the case of a two-option moral deliberation there are perhaps numerous *moral* options: (1) to try as hard as possible to make oneself intend to do the right thing (deploying as much willpower toward behaving in that way, for the sake of its rightness, as would be necessary for the origination of an intention to behave in that way); (2) to try, but not as hard as possible, to make oneself intend to do the right thing (deploying insufficient willpower to make oneself want to behave in that way, when one has sufficient willpower for this, or deploying less than all of one's willpower, when one doesn't have sufficient willpower to make oneself behave in this way); and (3) to refrain from deploying any willpower at all toward behaving in the right way. The second moral option, in which we don't try as hard as we might have, obviously splits into an infinite number of sub-options, in which we try to varying degrees (deploying varying percentages of our willpower) to behave in a certain way. One couldn't know how much effort would be required for the origination of an intention to do the right thing.

Point 8. Our heredity and experiences initially determine the quantity of our willpower as well as the range and force of our desires. Our making ourselves want to do whatever is right, or want this more strongly, might then over time alter both the quantity of our willpower and the range and force of our desires. It might not only leave us with an enduring, or enduringly strengthened, if slowly dwindling, desire to do whatever is right. It might also strengthen other desires, or lead to the growth of new ones, by

habituating us to behave in certain ways that we then come to want for their own sake. However, it would do all of this only to a limited extent. Someone whose genes and early experiences make him a violent, indolent person might be able to make himself less violent, and more industrious, but he might not be able to make himself truly civil and productive. (While our heredity and experiences limit how much willpower we can have, they do *not* determine whether or not we deploy it, nor do they determine how much willpower we deploy, both in two-option decisions, when we try, but not as hard as we might, to make ourselves do the right thing, and in multi-option decisions when the deployment of more or less willpower has definite behavioral consequences.)

The remainder of this chapter will address relevant issues.

(a) The specificity of moral improvement. Someone makes himself intend to do the right thing by making himself begin to want to do whatever is right, or by strengthening his existing desire to do whatever is right. Willpower may therefore be thought of as an ability to want to do whatever is right, or (somewhat misleadingly, since this suggests that desire is a sort of energy, rather than a state or continuing event) potential desire to do whatever is right. The following objection might then be made. According to point 2, above, when we originate a desire, or strengthen an existing one, in originating an intention to behave in a certain way, we modify our personality so we now tend to have such desires, or to have them at this new level of force. But the deployment of willpower is simply potential desire to do whatever is right. So, our origination of an intention to do the right thing should modify our personality in this way: we would now tend to desire to do whatever is right to a greater extent than we have in the past. But then someone who originates an intention to be more helpful than he was formerly

inclined to be should thereby become more likely not only to behave more helpfully in the future than he otherwise would have, but also to behave more honestly in the future than he otherwise would have, for it is also right to behave honestly rather than dishonestly. Yet it is easy to imagine someone making himself become a more helpful person through his moral decisions without thereby making himself become a more honest person. This suggests that even if willpower is necessarily rightness-oriented, it must be a potential desire to behave in some specifiable but as-yet-unspecified way in order to do whatever is right, so that when it is deployed it becomes an actual desire to behave helpfully in order to do whatever is right, or an actual desire to behave honestly in order to do what is right. Or, there must be distinct types of willpower, one being a potential desire to behave helpfully in order to do whatever is right, and another being a potential desire to behave honestly in order to do whatever is right.

My response to this objection is that there are at least two reasons why someone will be more likely to behave helpfully in the future, but *not* necessarily much more likely to behave honestly in the future, when he deploys willpower toward behaving helpfully. First, he may not think that being honest is in general more right than following one's dishonest inclinations, when it isn't obviously more helpful. Honesty may seem to him to be a mean, bourgeois sort of virtue. Second, his nonmoral desire to be helpful may already be so strong that his new desire to do whatever is right is all that is needed to make him want to behave helpfully, while his nonmoral desire to behave honestly, if there is such a thing, is too weak for his new desire to do whatever is right to make him want to behave honestly very often.

b. Apparent counter-examples. According to the account of pure moral responsibility that I am proposing, we either make ourselves intend to behave in the way that we think is more right than the way in which we are inclined to behave, or we follow our inclination instead. It might seem, though, that there are cases in which we do not merely refrain from making ourselves intend to do the right thing. We seem sometimes to make ourselves do something that we think is *less* right than what we were inclined to do. This would include cases in which it seems to us that we have had to make an effort to overcome a nonmoral revulsion, where the less revolting option seems to be more right, and cases in which it seems that we actually have to overcome our scruples, our desire to do whatever is right. Someone might make himself do grisly deeds that sicken and/or sadden him in order to attain wealth and power, while recognizing that this activity is wrong. Or, someone might have to make himself ignore the promptings of his guilty conscience in order to do these deeds; it seems that nonmoral revulsion alone wouldn't be enough to incline him not to do them. He overcomes his scruples. We can call these two characters "Macbeth I" and "Macbeth II", respectively.

I think that such cases can be plausibly explained in a way that is compatible with the proposed account of pure moral responsibility, even if we assume that Macbeths I and II really do originate the activity that they think they originate. Macbeth I doesn't really think that his activity is wrong. He just *calls* it wrong, because this is how most people label activity of this sort. On the contrary, he thinks that it is right, without realizing that he thinks that it is right. Macbeth II goes back and forth between thinking his activity right and thinking it wrong. He has a guilty conscience during the moments when he thinks it wrong, and makes himself perform it during the moments when he thinks it

right. Meanwhile, during these moments when he thinks it right, he feels a nonmoral revulsion against the activity. He confuses his recurring but not presently felt sense of guilt with this nonmoral revulsion. That is, he mistakes his nonmoral revulsion for a sense of guilt. This nonmoral revulsion really *is* enough to incline him not to perform the activity. His nonmoral revulsion might be a strong desire to avoid the sense of impending cosmic doom that his deeds induce in him, for example. The felt boundary between this anxious foreboding and his sense of guilt might be indistinct.

Someone might go back and forth between thinking his activity right and thinking it wrong because he focuses on different aspects of his activity at different moments. Grandiose delusion might make Macbeths I and II think that whatever advances their careers is truly right, and so when they focus on the career-advancing aspect of their activity they think it right. When Macbeth II focuses on the other-harming aspect of his activity, however, he thinks it wrong.

Similarly, someone might think that his adulterous affair is wrong when he focuses on its deceitfulness, while thinking it right when he focuses on its renewal of his vitality. Even when, in focusing upon its renewal of his vitality, he thinks it right, he might not realize that he thinks it right, because the label “right” isn’t usually applied to adultery. At the same time, he might have a strong non-moral desire not to continue the affair. He might take this nonmoral revulsion to be moral revulsion because it takes the form of some subtle but pervasive anxiety, just as in the case of Macbeth II, and because he expects to feel a moral revulsion that he does not, in fact, feel at this moment. Its subtlety permits him to see it as the moral revulsion that he expects. So, while he thinks

that he overcomes moral revulsion in order to attain a nonmoral end, he in fact overcomes nonmoral revulsion in order to attain a moral end.

c. Aiming at lesser wrongness. It often seems to us that we make ourselves intend to behave in a certain way that primarily presents itself to us as *less wrong*, rather than as more right. We seem to do so for the sake of its being less wrong that we behave in this way, rather than for the sake of its being more right than we behave in this way. Apparently, we make ourselves intend to do “the less wrong thing,” not the right thing. For example, we may choose to insult rather than bludgeon someone who angers us. Insulting him doesn’t seem to us to be at all right, and so it can’t be said that we choose to insult rather than bludgeon him for the sake of the greater rightness of insulting him. Rather, it just doesn’t seem to us to be as wrong a thing to do as bludgeoning him would be. I am inclined to resist this bivalent picture of moral motivation, in which the appearance of lesser wrongness and the appearance of greater rightness are both positive motivating factors. A univalent picture, in which the appearance of greater rightness is the only moral motivator, is more attractive because it is simpler. Also, we commonly think that wrongness is only a deficit of rightness, so that if we recognize something as less wrong, we do so only insofar as we recognize it on a more basic level as being more right. So, I suggest that when we seem to make ourselves behave in a certain way (such as insulting someone) in order to do what is less wrong than what we are inclined to do (such as bludgeoning him), what is going on is this. We simultaneously see this behavior as (1) wrong in itself (for insulting him is wrong) and as (2) an instance of a general mode of behavior that is further along toward the rightness-end of the relevant general behavioral axis. In the insults-vs.-bludgeoning case, the relevant general behavioral axis

has loving-kindness at one end and murderous violence at the other end. In making ourselves intend to insult rather than bludgeon someone, then, we make ourselves intend the insulting behavior for the sake of its being further along than bludgeoning behavior toward the extreme of loving-kindness. Thus, we make ourselves intend it for the sake of its greater rightness rather than for the sake of its lesser wrongness. But at the same time, we do see it in relation to the wrongness-end of the kindness-to-violence behavioral axis. Seeing it in relation to the wrongness-end of this behavioral axis, we see it as being in itself wrong, only less wrong than bludgeoning the offender would be. Its appearance as less wrong, though, is not what motivates us to make ourselves intend it. We make ourselves intend it for the sake of what seems to be its relatively greater proximity to the *rightness*-end of the behavioral axis.

d. Unsuccessful originative attempts. The proposal that someone can try unsuccessfully to make himself intend to do so something may seem doubtful. Indeed, I think that it is doubtful. Still, it seems to me to be preferable to the alternative – that we either succeed in making ourselves intend to do something, or we don't try at all. Someone's unsuccessfully attempting to make himself intend to do something would consist in his deployment of an insufficient quantity of willpower toward behaving in the way that seems most right to him. This would occur in cases of two-option moral deliberating. I have noted that true two-option moral deliberation is rare. Still, it is sometimes necessary to decide upon one of only two morally significant options. We might have to decide, for example, whether or not to have a drink or take a bribe. In such cases, it might seem implausible that someone could deploy all of his willpower toward doing what he believes to be right, yet have insufficient willpower to originate an

intention to do it. The difficulty here lies in finding some empirical evidence of a distinction between cases in which we don't deploy any willpower at all toward doing what we think right, and therefore end up following our inclination, and cases in which we do deploy willpower, but don't have enough of it to keep ourselves from following our inclination. This difficulty might motivate us to endorse a proposal that offers an easy escape from it: the proposal that in two-option cases where the total quantity of someone's willpower is insufficient to overcome the force of his contrary desires, he simply cannot deploy any willpower at all. He would be morally paralyzed, and his existing desires would determine his intentions. It isn't as though, recognizing that he doesn't have enough willpower to overcome his contrary desires, he *refrains* from deploying his insufficient willpower in a spirit of futile heroism. Rather, we simply *can't* deploy any willpower when our willpower is insufficient to overcome the desires that incline us against doing the right thing. We can't even try to deploy our willpower.

But this proposal assumes that some sort of willpower-suppressing mechanism is activated by the presence of an intention to do what is thought to be wrong, when someone doesn't have enough willpower to alter the configuration of his desires. The existence of such a mechanism is certainly not empirically evident. The addition of such hidden mechanisms to our psychic picture should be resisted, as an inelegant complication. Rather than adding this mechanism to our psychic picture, we should therefore try to find some even moderately plausible candidate for empirical evidence of a distinction between cases in which we don't deploy any willpower at all toward doing the right thing and cases in which we deploy insufficient willpower. Endorsement of the proposal that there is such a distinction, supported by empirical evidence that is only

moderately plausible, is preferable to endorsement of the proposal, both inelegant and without any empirical support, that there is a willpower-suppressing mechanism in someone that is activated when he doesn't have enough willpower to make himself intend to do the right thing.

We can say that in cases of two-option moral deliberation, someone's unsuccessful struggle to do the right thing simply manifests itself in the *feeling* of struggling against inclination. But what mental events would make him feel that he is struggling in this way? Perhaps he repeatedly averts his attention from the objects of his predominant desires, turning it toward the rightness of the other way of behaving. He feels that this redirection of his attention is itself something that he intends to do for the sake of its rightness. So, he feels that he originates these redirections of his attention. But they are intermittent and short-lived, while for the most part he follows his inclination, in an at least implicitly intentional way. These are cases, it seems, in which someone never really *decides* to follow his inclination. He just keeps finding himself once more intending to do this thing.

This picture is most plausible in cases where the behavior in question is something that is so easy to perform that someone can begin to perform it without realizing that he is doing so. Then, once he realizes that he is doing it, he wants and is inclined to *keep* doing it, and intends to keep doing it. Only at this point does it become an act. The elaboration of a sexual fantasy might be an example. Someone repeatedly diverts his attention from his fantasizing, because he thinks that this is the wrong sort of mental activity to be engaged in, and keeps finding himself in the midst of it again, and finding himself once more in the midst of it he becomes inclined to persist. This picture

of moral struggle as grounded in aversions of attention is plausible not only for cases like this (in which someone struggles to make himself do something other than what he is already doing), but also for cases in which he struggles not to begin doing something new. So, a man tempted into adultery again and again reminds himself of his responsibilities to his family, of the debt of gratitude that he owes to his wife, and so forth, but his attention again and again returns to the woman before him, and each time that his attention returns to her his adulterous intention is reestablished.

In the case of an adulterous inclination, luck influences the outcome, for after a while the act might no longer be an option. Where Rick and Phil both have insufficient willpower to overcome their adulterous desire, but the insufficiency is not quite so great for Rick as it is for Phil, then Rick would be able to divert his attention from the idea of the act more frequently, and for longer periods of time, than Phil would; Rick might thereby slow his progress into infidelity until it is no longer an option. Similarly, an insulted man again and again reminds himself of his responsibilities to his family, responsibilities that he will not be able to fulfill if he is imprisoned (the likely consequence if he strikes down the one who has insulted him) – considerations motivating him to moderate the expression of his anger by deploying his willpower. But again and again his attention returns to the insult.

Shall we say, then, that a moment arises at which the insulted man's violent desire completely seizes him, and he swings the bat? If so, then, at that moment he really can't deploy any willpower toward doing the right, or more-right, thing. Perhaps this is because the possibility of doing anything other than swinging the bat doesn't even occur to him. Maybe the recurrent thought that alternate behavior is possible opens a window

of opportunity for the deployment of willpower, and when this thought is absent it doesn't matter how much willpower one has; one will follow one's inclination. So, everything depends upon seizing the opportunity to make oneself do the right thing when the thought of doing the right thing occurs.

It might be objected that these redirections of attention from what we most want to the idea of doing the right thing are really only evidence either of (1) a sustained deliberation that precedes deciding, or (2) the repeated momentary deployment of a quantity of willpower that would be sufficient to get the prospective adulterer home to his wife and children if he didn't keep changing his mind and ceasing the deployment. After deliberating over whether or not to deploy any willpower in opposition to his bad desire, the insulted man finally decides not to; or else, after deploying sufficient willpower to prevent any violent action for a while, he stops deploying any willpower at all and gives full expression and full satisfaction to his bad desire. I grant that this objection has some force; however, all that we require is *moderately* plausible evidence of deployment of insufficient willpower in a case of two-option moral deliberation.

A case of long-term multi-option deciding might be composed of a series of short-term two-option cases. Thus, someone might pursue an adulterous relationship with a greater or lesser degree of intensity, by copulating with his mistress on a greater number of occasions, each one of which is a two-option case. Or, someone can get more or less drunk by having a greater number of drinks, where the having of each drink is a two-option case. The strength of someone's bad desire and the quantity of his willpower both presumably fluctuate over time. Thus, in struggling against his adulterous desire every time that adulterous sex is an option for him, forcefully directing his attention away from

his lover and toward mental images of himself spending “quality time” with his wife and children, an adulterer might succeed in spending only half as many nights with his lover as he would have had he simply given himself up to his desire. Here I don’t think that we have a single act of insufficient deployment of willpower toward a single large-scale, long-term action (the adulterous affair). Rather, there is a series of deployments, some sufficient and some insufficient to keep the man away from his lover’s bed.

e. Weak will. I hold that we can deploy some or all of our willpower toward behaving in a certain way in order to do what is right even when our deployment of all of our willpower would be insufficient to make us intend to behave in that way. Our willpower would be insufficient because our opposing desires are too strong for our limited quantity of willpower to overcome. This proposal offers us a way to understand cases of “weak will.” The typical case of weak will is that of the drunkard who “tries” to stop drinking, but can’t, because his craving for alcohol is so strong. According to the ordinary understanding of what it is to be morally weak-willed, the weak-willed person is one who intends to do the right thing but can’t because his opposing desires are too strong. But if someone’s intention to do something is the result of the configuration of his desires, as I hold, then this ordinary understanding of what it is to be weak-willed must be rejected. Someone who intends to do something cannot possibly be prevented from doing it by the strength of his opposing desires. If these opposing desires are strong enough to prevent him from doing it, he cannot be understood as intending to do it.

As an alternate understanding of what it is to be weak-willed, then, I propose that the weak-willed person is one who deploys all of his willpower toward doing something, when this total deployment of his willpower is still insufficient to overcome the force of

his opposing desires. Someone's weak-willedness is supposed to lessen the quantity of blame that he deserves for doing the wrong thing. According to my proposal, his being weak-willed would not lessen the quantity of ordinary blame that he deserves for doing the wrong thing, but might be incompatible with his deserving any pure blame for doing the wrong thing. Someone's deploying all of his insufficient quantity of willpower toward behaving in a certain way would contrast with someone's having enough willpower to overcome his opposing desires, but not deploying enough of it to do so, and someone's not having enough willpower to overcome his opposing desires and not deploying as much of this insufficient quantity as he might. Someone who has enough willpower to overcome his opposing desires but does not deploy enough of it to do so is certainly purely morally responsible for his wrong behavior, and could not be regarded as weak-willed. Someone who doesn't deploy all of his insufficient willpower toward doing the right thing looks more like a weak-willed person, because his willpower would be insufficient to get the job done even if he deployed all of it. But he still seems to deserve some pure blame for not trying as hard as he might have to intend to do the right thing, and so does not seem to be what is meant by talk of weak-willed people.

It might be urged that we make an exception in the case of someone who has deployed all of his willpower unsuccessfully toward behaving in the right way on several previous occasions, and who therefore expects that future deployments of all of his willpower will also be unsuccessful, and so no longer deploys any willpower toward behaving in that way. (For example someone has tried to stop drinking before, and has concluded that the effort is useless.) I would say, though, that nobody can be sure that his willpower is still insufficient to get the job done, and that persistence might even build up

the quantity of someone's willpower, in the way that someone's daily unsuccessful efforts to lift a heavy object might build up his muscles to the point that he becomes able to lift the object. So, I would still not call this person weak-willed, in the ordinary, exculpating sense of the expression. (Obviously, though, he can be said to be weak-willed in the same way in which someone with small, flaccid muscles can be said to be physically weak even though he doesn't even try to get anything done with them.) I would reserve that description for the person who actually tries with all of his might (deploying all of his willpower), albeit unsuccessfully, to do the right thing.

f. Moral compromises. In this section I consider cases in which someone deploys enough willpower to do what he thinks more right than what he is inclined to do, but not as much as would be required to do what he thinks even *more* right, although he could deploy additional willpower toward doing this. This kind of scenario is distinct from those in which someone tries unsuccessfully to make himself intend to do the right thing. Here, there is only a *successful* attempt to intend to do something more right than the following of inclination.

This occurs in cases of multi-option moral deliberation. Suppose someone can do A, B, or C. He thinks that A is more right than B, and that B is more right than C. So, he thinks that A is most right. However, he is more inclined to do C than B, and more inclined to do C than A. The simplest explanation for his being in this situation would be that he desires only A, B, and C; that his having one of these is incompatible with his having any of the others; that there is no compromise solution that would partially satisfy A and B, or B and C, or A, B, and C together; and that he desires C more than he desires B and B more than he desires A.

Now, suppose that he has enough willpower to make himself intend to do A if he deploys all of his willpower. He might still deploy only enough willpower to make himself intend to do B. Introspective evidence suggests that this is so. We can feel guilty about having not done what seemed, and still seems, to be the most right thing that we could have done, while at the same time feeling some relief at the thought that at least what we did was more right than what we were inclined to do. Our feeling of guilt about not having done the most right thing suggests that we think we could have done the thing in the strongest sense of “could” – that we were psychically, as well as physically, able to have done it. Our being psychically able to do something that we aren’t inclined to do requires our having enough willpower to make ourselves intend to do it. On the other hand, our relief at the thought that at least what we did was more right than what we were inclined to do is in direct opposition to our sense of guilt, moderating the latter. So, it seems that our withholding of available willpower (what we feel guilty about) was not total – we did deploy some willpower, although not enough to do what we took to be the most right thing that we could have done.

Here two conceptual difficulties arise. First, it is easy enough to *say* that the explanation for someone’s making himself intend to do B, when it seems to him less right than A, which he can also make himself intend to do, is that he is more inclined toward B than toward A. But given that he would be making himself intend to do B for the sake of its being more right than C, why wouldn’t he just go all the way and make himself intend to do A for the sake of its even *greater* rightness? Why does he make rightness rather than the overall satisfaction of his existing desires his goal only to this extent?

This difficulty is really the same as the difficulty that we have in explaining why someone would make rightness rather than the overall satisfaction of his existing desires his goal, or vice-versa, when he has only two options – as when he decides to accept or reject a bribe. It is just as hard to see why someone should make rightness his sole goal, when he doesn't have to, as it is to see why he should make rightness a partial goal (or a more important goal than it has previously been for him), alongside the overall satisfaction of his existing desires. I think that while *we* cannot understand this, *some* possible being (God?) must be able to understand it. This is because if it were in principle incomprehensible then the deployment of willpower toward doing the right thing would be a purely random phenomenon, and as such could not constitute someone's origination of anything. *Our* incomprehension is perhaps due to the fact that the only explanations we can grasp are explanations appealing to efficient causes. A sufficient explanation through efficient causes of someone's doing or intending to do the right thing rather than following his inclination would be incompatible with his origination of this act or intention.

The second difficulty is this. Willpower is supposed to be potential desire for whatever is right. How, then, can we deploy it in making ourselves intend to do what is less than completely right, for the sake of the *greater* rightness of what we are making ourselves do, as compared to the following of inclination? Here it should be noted that this difficulty isn't unique to desire for whatever is right. Every basic desire must have an unlimited object: desire for beauty is desire for unlimited beauty, and desire for comfort is desire for unlimited comfort. So, when someone chooses household furnishings that are less beautiful than those which he might have chosen, because they

are more comfortable, but more beautiful than more comfortable alternatives, his operative desires are for unlimited beauty and comfort. Nevertheless, he chooses the furnishings for the sake of their being *more* beautiful than the more comfortable alternatives and for the sake of their being *more* comfortable than the more beautiful alternatives. Here, though, we have a simple balancing of prior desires for beauty and comfort. The parallel solution to the difficulty that we face is obvious but unsatisfying: when someone makes himself intend to do B rather than A or C (where A seems more right and C less right, but he is more inclined toward C and less inclined toward A), he makes himself intend to do B for the sake of its being more right than C *and* more in the line of his inclination than A. This solution is unsatisfying because we still don't understand why he doesn't instead make himself intend to do A for the sake of its simple rightness, or C for the sake of its being what he is most inclined to do. But, again, we *can't* understand this, although *some* conceivable being must be able to.

g. Natural saintliness. Here is an unattractive apparent consequence of the proposal that someone can only make himself intend to do the right thing. Imagine someone who not only always has reasonable beliefs about the rightness or wrongness of his possible acts but is also gifted with a perfect natural conscientiousness. His tendency to desire whatever is right is so pronounced that he is always inclined to do what he reasonably believes to be right. He therefore always intends to do what he reasonably believes to be right, without ever having to originate these intentions. Evidently, such a person, who may be called a "natural saint", is morally superior to someone who doesn't always naturally intend to do what he reasonably believes to be right. Yet the natural saint apparently can't originate any intentions, while the latter, ordinary person can.

Surely, someone who is able to originate intentions is superior to the natural saint with respect to his possession of this originaive power, even though he is inferior to him with respect to his moral character. This is disconcerting, because we would like to think that one person can be superior to another in *all* respects: we want a single, coherent ideal standard that we can hope to realize or have realized (if only by divine intervention) in ourselves.

To see whether or not this objection is sustainable, we first have to distinguish between beings who necessarily intend to do the right thing and beings who don't necessarily intend to do the right thing and then further distinguish, with respect to beings who necessarily intend to do the right thing, between those whose existence is necessary and those whose existence is not necessary. That gives us three categories of beings. (Beings who don't necessarily intend to do the right thing might also be divided into these two categories, those who necessarily exist, and those who, like us, don't necessarily exist. That would give us a total of four categories. However, this further division of beings who don't necessarily intend to do the right thing into those who necessarily exist and those who don't is irrelevant to our present purpose. The objection may be answered in the same way, whether or not the existence of beings who don't necessarily intend to do the right thing is necessary.)

Beings who necessarily intend to do the right thing would stop being themselves if they stopped intending to do the right thing, while beings who don't necessarily intend to do the right thing would still be themselves even if they stopped intending to do the right thing. The God of many theologians would be the one being that both necessarily exists and necessarily intends to do the right thing. Angels who cannot conceivably rebel

would be beings that don't necessarily exist but that do necessarily intend to do the right thing. And we, of course, don't necessarily intend to do the right thing. We might sometimes, or even always, intend to do the right thing, but don't *necessarily* intend to do the right thing. We would still be ourselves even if we stopped intending to do the right thing. Even if one of us were genetically programmed to always intend to do the right thing he could stop intending to do the right thing as a result of some kind of organic malfunction without ceasing to be himself. (Similarly, someone who is genetically programmed to be very strong can become weak as a result of a virus, and pills can mellow someone who is genetically programmed to be aggressive, without these people ceasing to be themselves.)¹⁰²

The objection fails with respect to beings like us, who don't necessarily intend to do the right thing. Even if such a being always intends to do the right thing he can be able to make himself intend to do the right thing. He just doesn't have the opportunity to use this ability. He is in exactly the same position as someone who presently intends to do the right thing, even though he hasn't always intended to do so and won't always intend to do so in the future. Someone isn't unable to make himself do the right thing just because he presently intends to do the right thing. To say that when someone who doesn't necessarily intend to do the right thing intends to do the right thing he is therefore unable to *make* himself intend to do the right thing is like saying that an army can't fight

¹⁰² Although Galen Strawson does not think that self-determination would have to be rightness-oriented if there were such a thing, his "Natural Epictetans" resemble my angels, who, according to my own moral self-determinism, would be unable to originate anything because they would necessarily intend to do whatever is right. Strawson's Natural Epictetans are "[A] race of gifted, active creatures who ... are never undecided in any way," so that "[it] seems likely that the notion of freedom will have very little content for them" (*Freedom and Belief* 249). I am assuming that the Natural Epictetans are "never undecided in any way" because they are necessarily decided. If it just happens to be the case, but doesn't have to be the case, that Strawson's Natural Epictetans are never undecided in any way, then they would be like my human saints, whose self-determining power *would* have "content" for them, since they might conceivably become inclined to do something other than the right thing.

when it is not at war. So, someone who always intends to do the right thing isn't therefore unable to make himself do the right thing. If he were to stop intending to do the right thing, then he would have an opportunity to use his ability to make himself intend to do the right thing. It doesn't make any difference whether or not he will *ever* have an opportunity to make himself intend to do the right thing. He is still able to do so, at least as long as it is *possible* that he not intend to do the right thing, because it is possible that he become inclined to do something else.

What about beings that necessarily intend to do the right thing, then? Here the necessity or non-necessity of such a being's existence would make a big difference. Even if a necessarily existing being necessarily intends to do the right thing (and here we are talking about the God of many theologians), he might still make himself intend to do the right thing. Only, he would *necessarily* make himself intend to do the right thing. For, he might necessarily exist *as* something that makes itself intend to do the right thing. Beings that necessarily intend to do the right thing but *don't* necessarily exist (angels that cannot conceivably rebel) are in a very different situation. Such beings could not have their necessitated intention to do the right thing as a result of their necessarily *making themselves* intend to do the right thing, since they are made as what they are by something other than themselves. It is this other thing, the thing that makes them as what they are, that makes them intend to do the right thing.

A major difference between our originaive ability and God's might be that while we can deploy our ability to make ourselves intend to do the right thing in cases where the right thing is not what we are inclined to do, God simply makes himself intend to do the right thing. (Or, might God necessarily have desires that he necessarily overcomes by

making himself intend to do the right thing?) But in any case, our ability to make ourselves intend to do what we think more right than what we are inclined to do would be a specific form of a more general ability to make ourselves intend to do the right thing that we would share with God and the angels. The proposal that someone can make himself intend to do the right thing turns out not to have unattractive consequences for us or for God. It only has unattractive consequences for the angels. It seems that the human natural saint can be superior to the angels with respect to his originative ability without being inferior to them with respect to his moral character.

The angels might retort that they are still better than the human natural saint with respect to the necessity of their intention to do the right thing, for with respect to this it is better to necessarily intend to do the right thing than not to necessarily intend to do the right thing. They might then go even further, and urge that someone who necessarily intends to do the right thing but can't make himself intend to do the right thing is better overall than someone who doesn't necessarily intend to do the right thing but can make himself intend to do the right thing. It seems to me, though, that what makes the necessity of someone's intention to do the right thing good is only that he will therefore intend to do the right thing. So, someone isn't made better by *necessarily* intending to do the right thing; rather what makes him better is just that he intends to do the right thing. It might be true that someone who intends to do the right thing but can't make himself intend to do anything is better than someone who doesn't intend to do the right thing but can make himself intend to do the right thing. However, someone who intends to do the right thing but can't make himself intend to do anything wouldn't be as good as someone who intends to do the right thing and can make himself intend to do something. The

necessity of the angelic intention to do the right thing would be irrelevant. So, the angels would be better than those of us who don't always intend to do the right thing but not as good as those of us who are natural saints, always intending to do the right thing.

One reply that self-defensive angels might make is surely incorrect. This would be the claim that the ability to make oneself intend to do the right thing is valuable only as a means to one's intending to do the right thing, a means that the angels simply don't need, since they already necessarily intend to do the right thing. It seems to me that, on the contrary, there is something irreducibly grand (I don't know how else to say it) about being able to originate intentions and acts.

h. Originating a way of life. It seems implausible that someone could have a general intention to do whatever is right. At most, he could hopefully imagine himself as always doing whatever is right. And, if someone can't have a general intention to do whatever is right he can hardly originate such an intention. But it seems that someone *can* intend to live in a certain way for the sake of the rightness of living in this way, where a "way of living" is something with more specific content than living conscientiously. Someone can intend to live a life that is primarily devoted to contemplation, or political service, or artistic creation, for the sake of what he takes to be the rightness of this way of living. (There is nothing relativistic about the claim that the different skills and interests of different people would make different ways of life right for them.) If a decision to live in a certain way is possible, then an intention to live in that way is possible, and, indeed, there are many examples of people deciding to live in certain ways. Moreover, such decisions seem at least sometimes to emerge from moral deliberations over whether to live as one is inclined to live (usually, one's present way of

living) or whether to live in the way that seems be right, or in any case in a way that seems to be more right. So, if we are going to accept the claim that people can originate intentions to do what seems to them to be more right than what they are inclined to do, we should accept the claim that people can make themselves intend to live in a way that seems to them to be more right than the way in which they are inclined to live.

Presumably, someone's decision to live in a certain way would entail, but not be limited to, his decision that when, at future times, he must choose from among different kinds of short-term acts, he will choose those that best fit or are most conducive to that way of living. So, in originating an intention to live in a certain way for the sake of its rightness he would be originating an intention to decide to do whatever best fits or is most conducive to his living in that way, given that these short-term acts are themselves right. Someone who originates an intention to live contemplatively will originate an intention to decide to read, write, or think at times when he doesn't have a more pressing duty to pay a bill or help someone. But his originated intention to live in this way for the sake of its rightness cannot be an intention to *originate* these particular decisions. For, in originating his decision to live in this way he would be intensifying his desire to do whatever is right. This same intensified desire to do whatever is right would be what tips the balance of his desires so that he will make particular decisions befitting the way of life that he has chosen. So, someone must think of his general originated intention to live contemplatively as itself ensuring, or making very likely, future decisions to read, write, or think, rather than partying, watching television, or joining political committees. This thought seems reasonable. Someone's origination of an intention to do something for the sake of its rightness modifies his personality, giving him a greater tendency to want to do

whatever is right. In this case, he believes that what is right is his living in a certain way. Given the persistence of this belief, his greater tendency to want to do whatever is right will make it likely that in the future he will decide to do whatever fits or is conducive to his living in that way.

It seems that someone's origination of an intention to live in a certain way should make his belief that it is right that he live in this way more durable. Otherwise, his commitment to this way of living would be shakier than we take such commitments to be. However, his continuing belief that it is right that he live in this way isn't all that will ensure that he continues to make the particular decisions that fit and sustain such a life. Once he really gets involved in his new way of life, nonmoral desires will probably play a large role in attaching him to it. He will have acquired the habit of doing things typical of such a life.

To be genuine, an intention to live in a certain way probably requires subordinate intentions to behave in ways befitting such a life in the near future, where these ways of behaving would be thought of *as* befitting such a life. Someone's decision to live contemplatively would have to involve a decision to, for example, stop by the bookstore that afternoon to purchase Plato's dialogues. He would intend to do this *as* befitting a contemplative life, for the sake of the rightness of living contemplatively.

Bibliography

Ainslie, George. *Breakdown of Will*. Cambridge: Cambridge University Press, 2001.

Bennett, Jonathan. "The Conscience of Huckleberry Finn." In Landau, Russ Shafer, ed. *The Ethical Life: Fundamental Readings in Ethics and Moral Problems*, edited by Russ Shafer Landau, 289-302. New York: Oxford University Press, 2010. First published in *Philosophy* 49 (1974).

Bratman, Michael E. "Identification, Decision, and Treating as a Reason." In Ekstrom 2001, 107-125. First published in *Philosophical Topics*, vol. 24, no. 2 (1996), 1-18.

Campbell, Charles Arthur.

In Defence of Free Will, With Other Philosophical Essays. London: George Allen & Unwin Ltd., 1967.

Selfhood and Godhood. London: George Allen & Unwin Ltd., 1957.

Chisholm, Roderick M.

"Human Freedom and the Self." In Ekstrom 2001, 126-137. First published in The Lindley Lecture, 1964 (Department of Philosophy, University of Kansas), 3-15.

Person and Object: A Metaphysical Study. La Salle: Open Court Publishing Company, 1976.

Clarke, Randolph.

"Toward a Credible Agent-Causal Account of Free Will." *Nous* 27:2 (1993): 191-203.

"Agent Causation and Event Causation in the Production of Free Action." *Philosophical Topics* 24 (fall 1996): 19-48.

Davidson, Donald. *Essays on Actions & Events*. Clarendon Press: Oxford, 1980.

Davis, Lawrence H. *Theory of Action*. Prentice-Hall: Englewood Cliffs, 1978.

Ekstrom, Laura Waddell.

Ed., *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder: Westview Press, 2001.

Free Will: A Philosophical Study. Boulder: Westview Press, 2000.

Fischer, John Martin. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell, 1994.

Flanagan, Owen. *The Problem of the Soul: Two Visions of Mind and How to Reconcile Them*. New York: Basic Books, 2002.

Foot, Phillipa. "Killing and Letting Die." In *Contemporary Moral Problems*, 9th ed., edited by James E. White, Thompson Wadsworth 2009, 166-171. From *Abortion: Moral and Legal Perspectives*, edited by Joy L. Garfield and Patricia Hennessey (Amherst: The University of Massachusetts Press, 1984), 177-185.

Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." In Ekstrom 2001, 77-91. First published in *Journal of Philosophy* LXVIII, 1 (January 1971): 5-20.

Franklin, R.L. *Freewill and Determinism: A Study of Rival Conceptions of Man*. New York: Humanities Press, 1968.

Galavotti, Maria Carla. *Philosophical Introduction to Probability*. Stanford: CSLI Publications, 2005.

Gewirth, Alan. *Self-Fulfillment*. Princeton: Princeton University Press, 1998.

Ginet, Carl. *On Action*. Cambridge: Cambridge University Press, 1990.

Greenspan, P. S. "Behavior Control." In *Moral Responsibility*, edited by John Martin Fischer, 191-204. Ithaca: Cornell University Press, 1986. First published in *Philosophical Review*, 87 (April 1978), 225-40.

Hare, R. M. *Freedom and Reason*. Oxford: Oxford University Press, 1963.

Kane, Robert.

The Significance of Free Will. New York: Oxford University Press, 1996.

"Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism." In Ekstrom 2001, 158-180. First published in *Journal of Philosophy* XCVI, 5 (May 1999), 217-240.

Kenny, Anthony. *Will, Freedom and Power*. York: Harper & Row Publishers, Inc., 1976.

Levin, Michael.

"A Compatibilist Defense of Moral Responsibility" In *Introduction to Philosophy: Classical and Contemporary Readings*, 3rd edition, edited by Louis P. Pojman, 425-436. New York: Oxford University Press, 2004.

“Compatibilism and Special Relativity.” *The Journal of Philosophy*, Volume CIV, No. 9, September 2007: 433-463.

Lewis, C.S. *Mere Christianity*. New York: Simon & Schuster Inc., 1996.

Mele, Alfred. *Springs of Action: Understanding Intentional Behavior*. New York: Oxford University Press, 1992.

Mellor, D. H. *Probability: A Philosophical Introduction*. Abingdon: Routledge, 2005.

Mill, John Stuart.

On the Logic of the Moral Sciences (A System of Logic, Book VI). Edited by Henry M. Magid. Indianapolis: The Bobbs-Merrill Company, Inc., 1965.

Utilitarianism, On Liberty, Considerations on Representative Government. Edited by Geraint Williams. London: J. M. Dent 1993.

Milo, Ronald D. *Immorality*. Princeton: Princeton University Press, 1984.

Moore, G. E. *Principia Ethica*. Buffalo: Prometheus Books, 1988.

O'Connor, Timothy. *Persons and Causes: the Metaphysics of Free Will*. New York: Oxford University Press, 2000.

Plato.

Complete Works. Edited by John M. Cooper. Indianapolis: Hackett Publishing Company, 1997.

Symposium. Translated by Alexander Nehamas and Paul Woodruff. In *Complete Works*, 457-505.

Gorgias. Translated by Donald J. Zeyl. In *Complete Works*, 791-870.

Prichard, D. H. “Does Moral Philosophy Rest on a Mistake?” In *Contemporary Ethical Theory: A Book of Readings*, edited by Joseph Margolis, 47-66. New York: Random House, 1966. First published in *Mind*, XXI (1912), 487-499.

Pojman, Louis P. “Strengths and Weaknesses of Utilitarianism.” In *Exploring Ethics*, edited by Steven M. Cahn, 104-111. Oxford: Oxford University Press, 2009. First published in Louis P. Pojman, *How Should We Live? An Introduction to Ethics* (Wadsworth 2005).

Reid, Thomas. *Essays on the Active Powers of the Human Mind*. Cambridge: M.I.T. Press, 1969.

Ross, W. D. *The Right and the Good*. Edited by Philip Stratton-Lake. Oxford: Oxford University Press 1930, 2002.

Simpson, Peter.

Goodness and Nature: A Defense of Ethical Naturalism. Martinus Nijhoff Philosophical Library, Vol. 22.

Vices, Virtues, and Consequences. Catholic University of America Press, 2001.

Stocker, Michael. "Desiring the Bad: An Essay in Moral Psychology." *The Journal of Philosophy*, Vol. 76, No.12. (Dec., 1979): 738-753.

Strawson, Galen.

Freedom and Belief. New York: Oxford University Press, 1986.

"The Unhelpfulness of Indeterminism" (Review of Robert Kane, *The Significance of Free Will*). *Philosophy and Phenomenological Research*, Vol. 60, No. 1. (Jan., 2000): 149-155.

"Free Agents." *Philosophical Topics*, Vol., 32, Nos. 1& 2. (Spring & Fall 2004): 371-402.

Strawson, Peter. "Freedom and Resentment." In Ekstrom 2001, 183-204. First published in *Proceedings of the British Academy*, Vol. 84 (1962).

Taylor, Richard.

Action and Purpose. Englewood Cliffs: Prentice-Hall, Inc., 1966.

Metaphysics. Englewood Cliffs: Prentice-Hall, Inc., 1963.

Van Inwagen, Peter.

An Essay on Free Will. Oxford: Clarendon Press, 1983.

"Free Will Remains a Mystery" (The Eighth *Philosophical Perspectives* Lecture). In *Philosophical Perspectives, 14, Action and Freedom, 2000*, edited by James E. Tomberlin, 1-19. Malden: Blackwell Publishers, 2000.

Wallace, R. Jay. "Normativity and the Will." In *Agency and Action* (Royal Institute of Philosophy Supplement: 55), edited by John Hyman and Helen Steward, 195-216. Cambridge: Cambridge University Press, 2004.

Warner, Richard. *Freedom, Enjoyment, and Happiness: An Essay on Moral Psychology*. Ithaca: Cornell University Press, 1987.

Watson, Gary. "Free Agency." In Ekstrom 2001, 92-106. First published in *Journal of Philosophy* LXXII, 8 (April 1975): 205-220.

Wolf, Susan. *Freedom Within Reason*. New York: Oxford University Press, 1990.