

28. Daniel and Frances Howard Snyder, "How an Unsurpassable Being Can Create a Surpassable World," *Faith and Philosophy* 11 (1994): 260-268. Rowe replied in "The Problem of No Best World," *Faith and Philosophy* 11 (1994): 269-271, as well as in "Can God Be Free?"

29. "Can God Be Free?"

30. *The Christian God*, p. 135.

31. "The Problem of Divine Freedom," p. 255.

32. I believe this term is due to Daniel Dermet, but I have not been able to find the source.

33. Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

34. I do not pretend to be summarizing van Inwagen's arguments here, which are considerably more involved and ingenious than we need to see now.

35. Thomas Talbott defends a similar view in "On the Divine Nature and the Nature of Divine Freedom," *Faith and Philosophy* 5 (1988): 3-24. Talbott quotes the following instructive passage from C. S. Lewis:

Whatever human freedom means, Divine freedom cannot mean indeterminacy between alternatives and choice of one of them. Perfect goodness can never debate about the end to be obtained, and perfect wisdom cannot debate about the means most suited to achieve it. The freedom of God consists in the fact that no cause other than Himself produces His acts and no external obstacle impedes them—that His own goodness is the root from which they all grow and His own omnipotence the air in which they all flower." *The Problem of Pain* (New York: Macmillan, 1962), p. 35.

36. I am grateful to Richard Feldman and John Bennett for helpful discussion of the issues of this paper and to William Rowe for allowing me to see an advance copy of "Can God Be Free?"

LIBERTY OF THE HIGHER-ORDER WILL: FRANKFURT AND AUGUSTINE¹

John J. Davenport

In Augustine's early account of good and evil (in *On Free Choice of the Will*), and in his story of his own moral struggles (in the *Confessions*) we find a view similar to Harry Frankfurt's account of the first and second orders of the will. However, while Frankfurt thinks that his hierarchical account of the will provides evidence against the libertarian principle (PAP) that alternative possibilities are required for a person to be morally responsible for her actions, Augustine's account shows that this is not so. Rather, Augustine holds that moral responsibility for our character as constituted by our volitional identifications requires alternative possibilities of the higher-order will. Such 'liberty of identification' can be required even if we reject PAP as a condition on responsibility for outward acts. I explain this in terms of a tracing-defense of a restricted libertarianism based on a principle of responsibility for character, and I compare the resulting model with Robert Kane's conception of ultimate responsibility. After responding to Frankfurt's objection that the order-asymmetry in such a model is implausible or unmotivated, I argue that such a model is immune from traditional or non-global Frankfurt-style counterexamples to PAP-type principles.

In two famous articles, Harry Frankfurt developed and defended the compatibilist position that alternate possibilities of action are not required for moral responsibility.² Yet, as I will argue, Frankfurt's analysis of the will is closely related to St. Augustine's early work, which is the main classical source of the libertarian doctrine that a morally responsible agent must be able to do or will otherwise than she did. While Frankfurt's ideas about autonomy can help in understanding Augustine, Augustine's early position on good, evil, and responsibility for character also poses a unique challenge to Frankfurt's views on the freedom required for moral responsibility. This challenge has not been widely recognized in the burgeoning literature influenced by and responding to Frankfurt's insights,³ but it forms a powerful version of what I have recently called the "tracing defense" of alternative-possibility requirements for moral responsibility.⁴ Developing these implications shows that Augustine's early account offers a novel libertarian corrective to Frankfurt's model.

I. Frankfurt's Hierarchical Model of the Will

Harry Frankfurt's main debt to the Augustinian tradition is found in his volitionalist account of personhood, which was developed in response to



mid-20th-century theories that reduced personhood to a combination of consciousness, rational judgment, and embodiment. In response to Peter Strawson in particular, in his 1971 paper, "The Freedom of the Will and the Concept of a Person," Frankfurt argued that persons are distinguished from other animals by their capacity for what he calls "second-order volitions." He defined an agent's first-order will as "the desire (or desires) by which he is motivated in some action he performs or...the desires by which he will or would be motivated if or when he acts" (p.14).⁵ To clarify this, we may say that a first-order motive in general is a subjective mental state that inclines the agent to some act which does not have mental states of the agent herself as its end or goal. In other words, first-order motives are outwardly directed practical attitudes, aimed at ends external to the agent's psyche. In contrast, Frankfurt explains that a second-order volition is the will to act on one kind of first-order motive rather than another. As Frankfurt first formulated it, a second-order volition is the will to have a certain first-order desire be the effective motive if or when one acts (p.15). Thus if we think of this 'effective motive' more specifically as the maxim or intention in terms of which an agent free from self-deception would understand her own behavior as *an act*, then a second-order volition can also be understood as the will to be moved to one kind of intentional *aet* as opposed to others, or to adopt one kind of maxim for action. More generally, second-order volitions are reflexive in nature: they aim at an end or goal that concerns the agent's own first-order motivational states.

Frankfurt believes that such higher-order volitions explain the phenomenon which he famously calls *identifying* with a motive, or with a kind of character. "Identification," in its special Frankfurtian volitional sense, refers to the intrapersonal experience of striving to become one kind of person rather than another: it involves working on the *social self* or outward character that one presents to the world. The person who identifies with a motive or reason for acting in a particular way experiences this motive as her own, or as authored by her. As opposed to a "wanton" who (like non-human animals) "does not care about his will" (p.16), someone who is a "person" in the full sense has second-order volitions and is therefore not neutral or unconcerned about the will on which she acts. Rather, she *identifies* with one desire over others as the one she wants to be her first-order will. Frankfurt gives as an example an "unwilling addict" who "identifies himself, ...through the formation of a second-order volition" with the desire to refrain from taking drugs, even though he still acts on his craving for the drugs (p.18). Thus, on this account, identifying with a first-order motive, and in the process "alienating" other conflicting first-order desires on which one might act, both occur through the formation of second-order volitions.⁶

This analysis leads Frankfurt to two important conclusions. (1) In various circumstances, the first-order will that is constitutive of the act (i.e. the motive or maxim that gives the act its intentional meaning) may involve an intention with which we do not *identify* in the deeper personal sense, an intention that our inward self rejects in its second-order volitions. (2) And when we identify with the first-order motive on which we act, it need not be agent-caused for us to be morally responsible for our action. Frankfurt illustrates the second point by contrasting the unwilling addict with a

"willing addict." While the former is "helplessly violated by his own desires," and therefore presumably has diminished moral responsibility for his act (like someone acting intentionally but under duress), the latter willingly identifies with his irresistible craving for heroin, and is thus fully responsible for his action even though the addiction may have begun by accident, and now he cannot act otherwise:

His [first-order] will is outside his control, but by his second-order desire that his desire for the drug should be effective, he has made this will his own (p.25).

Thus identification with the first-order will *W*₁ that is causally operative in our behavior (making that behavior an action) is sufficient for moral responsibility for our action, even though the higher-order volition may not have caused the act or *brought it about* that *W*₁ was the motive on which we acted.

Frankfurt regards the willing addict and similar "over-determination" cases as counterexamples to the "principle of alternative possibilities" (PAP), different versions of which say (simplifying somewhat) that an agent is responsible for some *X* (a decision, intention, act or consequence) only if he had the power to bring about some significant alternative to *X*. Frankfurt's rejection of Augustinian *liberum arbitrium* as a condition for moral responsibility has led to a considerable and ongoing debate in recent literature focusing on whether Frankfurt-cases constitute real counterexamples to various possible versions of PAP, and if so, whether they also prove that responsibility is compatible with complete psychophysical determinism (in which all relevant future events are entailed by the laws of nature together with the past). But this debate has largely overlooked a crucial point that Augustine's ideas illustrate: the hierarchical analysis of volition by itself is still compatible with a libertarian account of the conditions of moral responsibility. The merits of this combination have not been fully explored, but deserve serious consideration.

II. Augustine's Hierarchical Conception of Good and Evil in the Will

Although he conceives the faculty somewhat differently than Frankfurt, Augustine also considers our will to be the distinguishing feature of human personhood, and he regards the agent's will as the center of her individual identity. In his *De Libero Arbitrio (On Free Choice of the Will)*, Augustine insists that the person's *seif* is most closely or authentically identified with his will: "If the will by which I choose or refuse things is not mine, then I don't know what I can call mine." Frankfurt's account of autonomy, which developed from his hierarchical analysis of the will, also requires that acts of volitional identification count as expressions of the agent's true self: unlike first-order motives, which can persist even when they are rejected by the higher-order will, the agent *eo ipso* identifies with the reflexive process or state of will that identifies him with a first-order motive. In other words, higher-order volitions are inalienable as long as they constitute identifications.⁷

Likewise, while Augustine never uses the language of second-order volitions, the notion is implicit in his early analysis of good and evil. In Book I of *On Free Choice of the Will*, Augustine focuses on the motives behind immoral actions, arguing that "evildoing" is always motivated by a perverse or "inordinate desire" for temporal or worldly goods whose possession depends partly on chance and fortune, i.e. "things that cannot be possessed without fear of losing them."⁹ In other words, Augustine holds that evildoing consists in acting on a certain unrestrained form of 'first-order will,' namely one that desires material and social goods to an inordinate degree, or even at the cost of unjust harm to others.¹⁰ Thus when Augustine argues later that "only the will and free choice can make the mind a companion of cupidity" or inordinate desires,¹¹ he is implying that we begin to act on such desires only because we freely accept or identify with them: whether we are reflectively aware of this or not, we will that an inordinate desire become the first-order intention on which we act. Augustine believes that we form higher-order volitions through which we actively help to shape our operative motives, accepting or even cultivating the motivational force that certain desires have for us, while resisting or working to break down the motivational force of other incentives.

This becomes particularly clear in the third book of *On Free Choice of the Will*, when Augustine's student Evodius asks "why that nature sinned which God foresaw would sin" and adds that he wants more than the explanation that it sins by free choice, since he is "looking for the cause of the will itself."¹² Augustine first responds that the search for such a cause would lead to an infinite regress, and so choice of one option over others cannot have any sufficient or determining cause outside the will itself.¹³ But Evodius's question can be understood not as asking for such an external explanation, but rather as seeking some account of the motive internal to the evil higher-order will, or the basis for the choice(s) through which inordinate desires become our dominant motives. Augustine eventually faces this deeper question when he considers the relation between will and reason. He says:

Only something that is seen can incite the will to act. We control whether we accept or reject whatever we see, but we do not control what we see. Therefore, we must acknowledge that the soul as a rational substance sees both superior and inferior things; from either class it chooses what it wills....¹⁴

The will must have objects or ends that it does not create. So it might seem that the evil will just chooses to pursue the lower goods as if they had an absolute value, even when pursuit of them conflicts with the higher goods. Yet why would anyone willingly choose what they perceive to be the lesser good, especially if they did not already desire it inordinately, as Augustine believes must have been the case with Adam before the fall, or with the Devil before his fall? At this level, what we need is some account of the motive or basis of the *higher-order* volition to desire the inferior goods unconditionally, or even at the cost of eternal goods of the soul.

Augustine's final answer is that the object of this evil higher-order volition is the agent's own self. Adam identifies with desire for the temporal

goods out of a "pride that turns one away from wisdom....And what is the source of this turning away, if not that someone whose good is God wants to be his own good, as if he were his own God."¹⁵ Likewise, Satan's rebellion involves willing to put himself in God's place, and what first suggests this to him is his mind's perception that he is a different being than God.¹⁶ Here it is clear that an evil will involves more than simply inordinate first-order desire for temporal goods. At least in first sin, it must also involve the higher-order volition to reject divine authority, or to overvalue lesser goods just because we want to determine for ourselves which goods are more or less valuable independently of God's will and law. Thus evil begins in rebellious obstinacy and pride in the higher-order will.¹⁷

The same notion of a higher-order volition standing above our given first-order motives is implicit in Augustine's later accounts of moral psychology, where he argues that memory, understanding, and will are capable of applying reflexively to one another and to themselves:

...For I remember that I have memory, and understanding, and will; and I understand that I understand, and will, and remember; and I will that I will, and remember, and understand....¹⁸

In *Free Choice of the Will*, the related claim that nothing "is so much in the power of the will as the will itself"¹⁹ occurs in the context of an argument that "it is up to our own will" whether or not we enjoy a "good will." Augustine defines a good will as one "by which we desire to live upright and honorable lives and to attain the highest wisdom."²⁰ Like the evil will, the good will so conceived involves higher-order volitions. It does not just consist in acting on ordinate first-order desires (which we might call 'good-doing' as the opposite of evil-doing). Rather, the good will involves the second-order volition that identifies with ordinateness of our proper first-order motives. The good person endorses and works to cultivate such motives because of their goodness. The inward state of moral uprightness or *rectitudo* of will in the Augustinian tradition—which was reinterpreted by Kant as "the motive of duty"—thus consists in a form of second-order volition that responds to a moral judgment about the worth of different sorts of first-order motive.²¹

These examples show that for Augustine, *moral worth* depends primarily on the state of our higher-order will: although everyone naturally has "the will to be happy" and thus desires goods that appear likely to contribute to their well-being, good or morally worthy persons are distinguished by "the will to live rightly,"²² or their rectitude in willing to act only on desires that are neither unjust nor inappropriate in the situation.²³ Likewise, an evil will requires volitional identification with perverse first-order motives; it is the second-order will in which "someone chooses to descend from the heights of wisdom and become a slave to inordinate desires."²⁴ Thus Augustine's conception of "will," as opposed to mere bodily action, is like higher-order volition in Frankfurt's sense: it is a reflexive commitment to being a particular kind of character, or to becoming a particular sort of person.²⁵

We see this even more clearly in Augustine's famous description of his struggles with himself in the *Confessions*. For example, in recounting his

desire to follow Victorinus's example and commit himself to the Christian God, Augustine writes:

...but I was held fast, not in fetters clamped upon me by another, but by my own will, which had the strength of iron chains. The enemy held my will in his power and from it he had made a chain and shackled me. For my will was perverse and lust had grown from it, and when I gave in to lust habit was born, and when I did not resist the habit it became a necessity...But the new will which had come to life in me and made me wish to serve you freely and enjoy you, my God, who are our only certain joy, was not yet strong enough to overcome the old, hardened as it was by the passage of time. So these two wills within me, one old, one new, one the servant of the flesh, the other of the spirit, were in conflict, and between them they tore my soul apart....In this warfare, I was on both sides, but I took the part of that which I approved in myself rather than the part of that which I disapproved. For my true self was no longer on the side of which I disapproved, since to a great extent, I was now its reluctant victim, rather than its willing tool.²⁶

In this passage, up to the italicized point, the description suggests two opposed forces of the same kind, or two states on a metaphysical par with one another, although the older one is more entrenched. But in the italicized portion this changes, and the description suggests that above these two desires stands a higher-order volition representing Augustine himself, by which he identifies with the kind of motive that he judges to be morally superior. Yet, as in some of Frankfurt's examples, this second-order volition does not immediately cause him to act on ordinate rather than concupiscent first-order desires. So, in language that clearly anticipates Frankfurt's, Augustine then describes himself as the "victim" of this habitual disposition that controls his acts, although he no longer identifies with it. As we see later, this is partly because his new second-order volition itself is not yet entirely "wholehearted," it is not yet a decisive commitment.²⁷ Eleonore Stump recognizes the same halfheartedness in Augustine's infamous prayer that God make him chaste, "but not yet." As she suggests, we can regard sincere prayer for a change in motive as an expression of the agent's higher-order volition. When such a prayer is less than wholehearted, it is because the higher-order will remains weak.

Since good and evil wills involve contrasting states of higher-order volition for Augustine, the freedom to turn from a good will towards and evil will, and *visa-versa*, ultimately resides in the power to form volitional identifications. When Augustine says that it is "up to us" whether we have a good will, this implies a requirement of alternative possibilities for the higher-order will itself. Thus he says, "If the movement of the will were not voluntary and under its own control, a person would not deserve praise for turning to higher things or blame for turning to lower things, as if swinging on the hinge of the will."²⁸ This *hinge* metaphor clearly indicates a libertarian freedom (in the proper alternative-possibilities sense)²⁹ which explains the human will's 'intermediate' nature between good and evil. Since the *inner character* of this will that turns us towards higher or lower things is

constituted by our higher-order volitions, our responsibility for the character of our will requires the liberty to form different volitional identifications or reflexive commitments themselves. The moral worth and/or culpability of "character" in this sense must be distinguished from mere praise or blame for particular outward actions. The good will is a great or infinite good according to Augustine's criteria precisely because its liberty saves the person's inner character from being determined by fate or fortune alone: their good second-order will "cannot be stolen or taken away from them against their will."³¹ This will cannot be determined by any "natural" necessity if it is to be culpable when it identifies with inordinate desires.

It is important to emphasize that such a libertarian account of volitional identification does not imply that the past has no effect on what options are open to the agent's higher-order will, or on how easily she may change her current volitional commitments. Just as our habits of first-order motivation may be tough to change, different possible higher-order volitions may be more or less *difficult* for an agent to establish in herself. Augustine's account does not require that the agent's alternatives in the higher-order will are *unlimited*, but it does imply that a higher-order volition—as exemplified in an upright will or an evil will—is never necessitated but rather remains one among a morally significant range of alternative possible identifications that are more or less accessible to us. Although the process through which one of these live options becomes actual is not *determined* by past choices and external causes,³³ it may still be inclined or conditioned by such factors.³⁴ Thus, while Augustine's account benefits from Frankfurtian clarification, it also requires its own unique and subtle version of libertarian freedom.

In sum, I have attributed to the early Augustine a version of libertarianism according to which *liberum arbitrium* in the higher-order will is essential to responsibility for our character first of all, and derivatively for our particular actions. This is a development of the traditional interpretation of Augustine's view. For example, explaining the distinction between *liberum arbitrium* and *libertas* or "higher freedom," Vernon Bourke writes:

The former is a part of man's created nature: the soul is endowed at birth with the ability to turn towards or away from its supreme good; this ability is *liberum arbitrium*. It implies a freedom of alternatives: to do what is good or what is evil. On the other hand, God may so dispose the human will that it inclines only toward its true good. This divine disposition would free the will from its tendency toward evil; such eminent freedom is *libertas*.³⁵

To this I have added Augustine's implicit recognition of *liberum arbitrium* in the higher-order will. However, both Augustine's understanding of grace and predestination, and his attempted answer to the dilemma of freedom and divine foreknowledge, naturally raise questions about whether it is right to attribute such a libertarian theory to Augustine. For example, David Hunt has argued that already in *On Free Choice of the Will*, Augustine endorses a conception of divine foreknowledge that makes our future choices temporally necessary, rejecting libertarian freedom as a requirement for moral responsibility.³⁶ Since the main goal of this article is

to assess the challenge that a hierarchical libertarianism poses for Frankfurt's account, for the moment I will simply assume that I have described Augustine's early view correctly. But I hope to address these important historical and exegetical questions in a planned follow-up article.

III. *Freedom of the Will and Liberty of Identification*

Like the early Augustine, in his first paper on the higher-order will, Frankfurt considered the liberty to form different second-order volitions or to change one's identifications to be part of "freedom of the will," which he distinguishes from "freedom of action." But unlike Augustine, Frankfurt never thought that moral responsibility requires a "free will" in his sense. Instead, he gives a detailed compatibilist account of freedom of action, which he thinks is sufficient for an agent to be responsible for a particular act. Roughly speaking, an agent exhibits "freedom of action" in Frankfurt's sense if he does what he wants (no impediment blocks his first-order motive from causing the intended act) and he does not have an opposing second-order volition (i.e. he does not identify with an opposite or incompatible first-order desire). By contrast, "When we ask whether a person's will is free we are not asking whether he is in a position to translate his first-order desires into actions" (p.20). While freedom of action concerns the relation of the first-order will to the behavior it tries to guide, "freedom of the will" concerns the relation of second-order volitions to the first-order motives with which they identify the agent. The primary condition for a free will is that her second-order volitions V_2 are successful, or bring it about that she acts upon the motives with which she identifies through V_2 (p.20). For instance, consider someone who acts out of nastiness towards a relative when he wills to act of out kindness instead (p.22). Because he is not moved by the sympathetic intention with which he identifies, "he finds himself a helpless or passive bystander to the forces that move him," and thus he does not have a free will (p.22).

Free will in Frankfurt's sense requires more than just that the agent's second-order volition agrees with his operative first-order desires, for in cases like the willing addict "their coincidence is not his own doing but only a happy chance" (p.20). Altering my example, suppose that just before he meets his relative, our man finds a \$100 bill; as a result, he is so happy that his urge to be cruel to his relative dissipates. He will surely still be somewhat disappointed with himself afterwards, since he knows that, but for his lucky find, he would have been nasty against his true will. In an agent exercising free will, the agreement between higher and lower-order will is not so dependent on luck.

Although he leaves this ambiguous in some passages, Frankfurt also implies that freedom of the will requires more than just a hypothetical relation between acts of volitional identification and the agent's first-order motives. Suppose that if agent S forms a higher-order volition V_2 identifying with desire D_1 , then it will successfully guide S to act on D_1 . This is not enough to ensure that S enjoys free will. If it were, then S could enjoy free will without actually forming any second-order volitions, as long as they

would be effective if formed. Then arguably even a wanton who lacked the mental capacity to identify with first-order motives could exhibit free will, as long as there was some logically possible world in which the wanton has enhanced mental capacities, would use them to form second-order volitions, and these volitions would be effective in guiding his first-order will.³⁷ But Frankfurt insists that a wanton can neither have nor lack freedom of will in his sense—it cannot even be an issue for the wanton (p.21). Thus freedom of the will requires more than hypothetical effectiveness of possible higher-order volitions. The agent with free will must also have it presently in her power to form higher-order volitions, a capacity that the wanton lacks. Thus Frankfurt says:

A person's will is free only if he is free to have the will he wants. This means that, with regard to any of his [first-order] desires, he is free either to make that desire his [first-order] will or to make some other desire his [first-order] will instead. Whatever his [first-order] will, then, the will of the person whose will is free *could have been otherwise*; he could have done otherwise than to constitute his [first-order] will as he did (p.24, my italics).

Frankfurt focuses here on the availability of alternatives at the first-order level, but the free-willed agent realizes these alternatives through her higher-order will. Her liberty to make either first-order desire D_1a or D_1b her will₁ must then derive from her capacity to form either the volition V_1a identifying with D_1a , or to form the volition V_1b identifying with D_1b .³⁸ Thus free will in Frankfurt's sense includes libertarian freedom to bring about different higher-order volitions, or to change our identifications.³⁹

Let us call this special condition 'liberty of the higher-order will' or the 'freedom to identify otherwise.' A person enjoys this liberty when, without requiring other changes in their present state as a precondition, he can form any one out of a significant range of possible higher-order volitions (or identifications). Note that this higher-order liberty could be enjoyed by *itself*, without the other condition involved in Frankfurtian freedom of the will. For example, although no crack-cocaine addict in the grip of compulsive desires enjoys Frankfurtian free will (since in the present, circumstances permitting, they will act on their craving for cocaine whether they identify with it or not) they could still enjoy liberty of the higher-order will. An addict enjoying such 'liberty of identification' is one who can make herself either a willing or an unwilling addict: although her will is not free, since she cannot actually refrain from acting on the addictive desire, she can either identify with this desire or alienate it (by committing herself to oppose it and trying to find ways to overcome it). Again, these alternatives may not be equally available to the agent, but she has more or less liberty in the higher-order will depending on the range of second-order volitions open to her, and how easy or difficult they may be to form.

Now this sort of scenario in which an agent enjoys liberty of identification but lacks full freedom of the will is familiar from Augustine's influential account of the bondage of sin. Admittedly, Augustine did not always clearly separate this liberty of the higher-order will from the ability to be

free from inordinate desire in our first-order will. Thus when he argues that "the mind must be more powerful than cupidity"⁴⁰ and so must be able to resist acting on inordinate desire, Augustine gives the appearance of clashing with Frankfurt's modern view that an unwilling addict, for example, can actually be coerced by physiological need to desire a drug. Yet in his own account of the volitional state of "difficulty" that results from first sin, Augustine in fact cites many similar experiences where we are "held back by some sort of necessity of carnal desire" which prevents us from acting on the very motives that we will to act on.⁴¹ To experience this as volitional difficulty, however, we have to be able to decide for ourselves not to consent to or identify with these inordinate desires: though our higher-order will may be too weak to control which desires we act on, we still control our higher-order will.

Because for Augustine we can be responsible not only for acquiring a good will but also for first sin—or the initial turning away from God that subjects us to later volitional difficulties—the range of alternative possible higher-order volitions that we can form must (at least initially) be a morally rich one. *Significant* freedom of the will must include the ability to identify with either ethically appropriate or unethical motives for acting. Thus Augustine argues that the human will is originally in a "state intermediate between wisdom and folly" and able to move itself in either direction;⁴² for this reason is the person's whole character subject to moral categories.⁴³ Moreover, Augustine takes on faith that it was by an original exercise of this capacity to sin that we lost a stronger power of the will:

But to accept falsehoods as truths, thus erring unwillingly; to struggle against the pain of carnal bondage and not be able to refrain from acts of inordinate desire: these do not belong to the nature that human beings were created with....[W]hen we speak of the free will to act rightly, we mean the will which human beings were created with.⁴⁴

In paradise, human beings would have had something like full freedom of the will in Frankfurt's sense, enabling them to exercise complete control over the first-order motives on which they act. But this ability is lost through an initial sin in the higher-order will itself. Nevertheless, every human being still retains the residual freedom, or *higher-order liberty*, to turn back to God,⁴⁵ or to identify with morally upright motives for action because they are ordinate—even if this inward *rectitudo* cannot effectively control our fallen first-order wills without divine assistance.

IV. Moral Responsibility for Character: Augustine's Challenge to Frankfurt

In light of these similarities between Augustine and Frankfurt, it becomes clear why Augustine's early account of the conditions of moral responsibility poses an important challenge to Frankfurt. Frankfurt holds that *neither* "free will" in his sense nor liberty of the higher-order will alone are necessary for responsibility for one's actions, while Augustine (at least in his earlier works) holds that liberty of the higher-order will is required. Like Frankfurt, Augustine traces responsibility for outward actions to responsibility for

higher-order volitions that identify us with the motives of those acts, but unlike Frankfurt, Augustine thinks that liberty of identification is necessary for us to be responsible for these higher-order volitions themselves.

As we saw, Augustine's account (as I interpreted it) implies that we have the liberty to identify either with good (or ordinate) desires or with inordinate desires as the motives of our actions. Identifying with the former is a state of moral virtue while identifying with inordinate desire is the original form of moral corruption, on Augustine's account. In other words, Augustine links liberty of identification in the higher-order will with the initial dual capacity for both *moral worth and unworth* in one's personality. Liberty of identification is an original condition of moral responsibility for one's *inward volitional character*, i.e. the character formed by long-term commitments or patterns in the higher-order will itself. This is to be distinguished from *outward character*, which consists in dispositions to act on various kinds of first-order motives, along with other familiar personality-traits. The unwilling alcoholic, for example, may have what psychologists call an 'addictive personality' in her outward character, but her inner character is formed in part by her commitment to self-reform.

According to Augustine, then, the moral worth of persons (as opposed to the isolated values of their particular actions) depends directly on the inward character made up of their enduring volitional identifications. The unwilling addict, for example, has at least a residual virtue that the wanton addict lacks.⁴⁶ But if the moral worth of our inner character is determined by our persistent higher-order volitions (or our enduring pattern of commitments), then the higher-order will itself must enjoy some measure of libertarian freedom in forming these volitions: our identifications can count as self-determined or as *our own* only if we could have identified otherwise. Of course, Frankfurt and his philosophical heirs have distinguished self-determination and libertarian freedom (in its proper sense as dual voluntary control over multiple alternatives). But the Augustinian model insists that these features converge at the highest level of the will, where it locates ultimate responsibility, even if they can diverge at the more derivative levels of human agency.

Frankfurt agrees with Augustine that one acquires a volitional *self* for which one is responsible precisely through authentic commitment to various desires or possible first-order motives. Thus the willing and unwilling addict differ in their moral worth because of the difference in the value of their volitional identifications. In a later paper, Frankfurt says that the person's self is constituted by the choices of higher-order will through which she tries to form a coherent and stable volitional character: by incorporating some preference-rankings and radically rejecting others, we "create a self out of the raw materials of inner life."⁴⁶ The early Augustine's position is similar. But because Frankfurt did not focus on the role identification plays in determining the self's *moral worth*, he did not initially ask if the agent had to be able to will otherwise when forming a higher-order volition for it to count as self-determined, or to constitute an authentic identification.⁴⁷ Yet in a series of more recent papers, Frankfurt has explicitly argued that we may be bound by what he calls "volitional necessities" in which we cannot avoid some of the higher-order cares and commitments

that define our true self.⁴⁸ In his view, we are still responsible for these decisive or wholehearted identifications, although it may have been inevitable that we would develop them. Furthermore, responsibility for these self-constituting commitments does not require the ability to change them even indirectly: it may rather be volitionally impossible for us even to question them, let alone to set out intentionally to change them. On this view, such necessary identifications still count as self-determined because they are expressions of our individual nature or personal essence.

By contrast, in *On Free Choice of the Will*, Augustine implicitly rejects any such notion of a personal essence: he insists that because our volitional identifications determine our moral worth, they must be generated through a process in which the agent could have willed₂ otherwise. Let us call this the principle of alternative possibilities for inner character:

PAP-C. Moral responsibility for one's inner character requires liberty of identification at the higher levels of the will, or in the capacity to form the volitional commitments that define our deep self: the agent is responsible for her higher-order volitions only if she could (at some point in her past) have voluntarily formed other volitions at these levels or voluntarily avoided her present identifications.

Although Augustine certainly offers no decisive proof of PAP-C, I think this principle has strong intuitive appeal: most people's considered convictions will agree with PAP-C unless they are given strong reasons to alter this intuition. But Frankfurt has not yet given us any such reasons to reject PAP-C, although he and his followers have challenged other PAP-type principles for actions, omissions, and their consequences.

The Tracing Defense of PAP-C. In general, the Frankfurtians have employed several famous over-determination examples to argue that agents can be responsible for something X (an action, its consequences, a prior intention, or perhaps a decision) when they have brought X about without interference, even though they could not have done otherwise.⁴⁹ I have argued in a recent review of Fischer and Ravizza's book that the intuitive appeal of all such cases depends on a distinction between two elements: "one (R) which explains why the agent is responsible for X, and another (I) which explains why X is inevitable," or why X could not have been avoided. In such cases, R intuitively makes the agent responsible for X even though X is inevitable, but *only* because we tacitly assume that the agent meets whatever conditions are required to be responsible for R itself. I would now add that, if it is to be plausible to everyone that the agent is responsible for X, then (1) the R and I factors must not only be separated, but (2) distinguished in such a way that it is open to the audience to take the R-factor as avoidable or evitable. As a result, libertarians can respond to such cases with what I call a tracing defense: they can say that in order to be responsible for R, the agent had to be able to avoid R, and thus responsibility for X still ultimately requires libertarian freedom. Any attempt to answer such a tracing defense with revised over-determination examples will require cases in which both the R and X-element are inevitable, and the libertarian

can hold that the agent is not responsible in such cases. For in such cases, the agent will be responsible not only in the actual sequence, but also in the counterfactual sequence, where R is caused by the external I-factor. Thus although Frankfurt-style examples may show that responsibility and inevitability can be locally compatible, for the libertarian, this is possible only because of the libertarian freedom which the agent enjoyed in bringing about the conditions that made her responsible for the X-element in the case, even though X is inevitable due to quite separate I-conditions.

We can apply this pattern of analysis to Frankfurt's willing heroin addict case. In this example, X is the taking of heroin, I is the physiological addiction or compulsive disorder that makes refraining from available heroin virtually impossible, and R is the agent's identification with his addictive craving, or his second-order will to remain an addict. The willing addict indeed seems responsible for taking his heroin, even though he could not do otherwise. This intuition depends on the plausible assumption that he is responsible for his second-order volition itself. Given the structure of this case, Frankfurt must make this assumption, but he cannot defend or analyze its conditions *within* the example: rather, it must be presupposed for the example to seem persuasive.⁵¹ Moreover, the plausibility of the assumption that the addict is responsible for his higher-order volition requires that we can regard it as avoidable. So if the willing addict seems responsible for his higher-order will, and thus for taking the drug, isn't that because we tacitly imagine this addict as having liberty of identification, and thus as having the volitional possibility of becoming an *unwilling* addict instead? Certainly the early Augustine thought that, with God's help, he could have worked this change in himself. I have found that undergraduates to whom I pose the case routinely imagine the addict as *freely* deciding to identify with his addictive desires rather than to alienate and fight them, and when questioned, they cite this presupposition in explaining their judgment that the addict's willingness makes him responsible for taking the drug. If instead we stipulate that the addict has been programmed to identify with and cultivate his cravings, or that he never had any other alternative higher-order volition open to him, many people will find that their initial inclination to hold him responsible for his addictive behavior evaporates. Because now the R and X factors are both inevitable, and we cannot trace the agent's responsibility for R to any prior condition R* which the agent did not inevitably satisfy, we are no longer sure that the agent is responsible. Rather, under these specifications, we doubt that his putative higher-order volition is really the agent's own, or truly self-determined, or really expresses the volitional identifications constitutive of a deep self.

Let me clarify that this tracing model is only meant to be a *defense* of PAP-C against familiar over-determination counterexamples to PAP-principles. It does not attempt to demonstrate on any independent grounds that PAP-C is correct. Rather, it only seeks to show that, given their structure, traditional over-determination cases leave libertarians free to hold PAP-C. I have not attempted here to answer objections to PAP-C that do not rely on Frankfurt-style cases,⁵² or to refute any rival compatibilist accounts of responsibility for our volitional identifications (for example, Frankfurt's neo-Leibnizian notion of personal essences, or Fischer and Ravizza's actual-

sequence account of "taking responsibility" for the psychological processes that explain our actions). Those antecedently disposed to prefer such semi-compatibilist accounts will have to ask themselves how much their disposition depends on the judgment that traditional Frankfurt-style cases are clear defeaters of PAP-type principles, for my argument does undermine this judgment. But I provide no direct defeaters of the semi-compatibilist theories here.⁵⁴ Rather, I have only argued that PAP-C cannot be defeated by traditional over-determination cases.

Comparison to Kane's Analysis. It is important to note that the extension of libertarian conditions on moral freedom by tracing principles has already been pioneered by Robert Kane. Kane's theory is a highly sophisticated development of Aristotle's earliest tracing theory: "if a man is responsible for wicked acts issuing from his character, then he must at some time in the past have been responsible for forming this character," where such responsibility requires the ability to do otherwise in choices that helped create this character.⁵⁵ In particular, Kane adds to the alternative-possibilities requirement what he calls the "condition of ultimate responsibility, or UR."⁵⁶ Kane's UR is a complex tracing version of PAP: the first part (R) defines "personal responsibility" for an event or state as requiring that the agent could have avoided something that causally contributed to the event or state, and the second part (U) says that

for every X and Y (where X and Y represent occurrences of events or states) if the agent is personally responsible for X, and if Y is an *arête* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y.⁵⁷

UR embodies one kind of libertarian tracing theory, because it entails "that some actions in an agent's life history must satisfy AP" (the condition of alternative possibilities), although not every free and responsible act must satisfy AP,⁵⁸ and its ultimacy condition requires that the agent be responsible for any actualized sufficient condition of any act imputable to him. The idea is clearly that responsibility for some X traces to responsibility for any actual state of affairs sufficient for X.⁵⁹

This is obviously somewhat different from the tracing defense I have sketched above, since my defense does not require that if the agent is responsible for some action A, then she is responsible for all actualized sufficient conditions for A, including the obtaining of any condition that makes A inevitable.⁶⁰ Rather, it traces responsibility for inevitable actions and decisions to states of the higher-order will, which Kane would include in our character. Thus my tracing approach can be used directly against traditional (non-global) Frankfurt-style cases, whereas Kane's principle is not tailored specifically for this purpose. Hence in *The Significance of Free Will*, Kane does not argue that in traditional counterfactual-intervener examples, the agent is responsible for an X which he also cannot avoid, yet only because this responsibility ultimately traces to a libertarian source. While his restricted libertarian theory of ultimate responsibility certainly encourages such a move, Kane sees that it does not as obviously answer

global Frankfurt-style cases. Instead he argues, like Carl Ginet and David Widerker, that in cases in which the actual sequence is indeterministic (so as not to beg the question against incompatibilists), the agent *can* avoid the X for which she is responsible. In particular, the indeterministic choices in which ultimate responsibility terminates cannot be rendered inevitable by a Frankfurt-controller.⁶¹

But this Ginet-Kane-Widerker approach has been challenged by several new cases, including Eleonore Stump's argument that "any mental act, even an act of willing, is correlated with a neural state which is not indivisible," although earlier portions of this neural sequence do not cause the act.⁶² Although Stewart Goetz and others have attempted to answer Stump, I think it may be better and more powerful to begin with the tracing defense against traditional Frankfurt-style cases, and then to see whether the sort of restricted libertarianism it implies can then shed light on what is wrong in the global cases.⁶³

However, the forms of restricted libertarianism consistent with my tracing defense will be similar to Kane's in many respects. In place of Kane's (U) they would have the requirement that to be responsible for X, we must be ultimately responsible for whatever conditions are sufficient for us to be responsible for X. In place of Kane's (R), they would have the thesis that ultimate responsibility for some state S, which is not traceable to prior responsibility for anything else, requires being able voluntarily to avoid S. When spelled out, these models will also entail Kane's conclusion that if ultimate responsibility must terminate in some state that is undeniably attributable to us, or by its very nature imputable to us, only a state originating from libertarian choice with dual or plural voluntary control can satisfy this condition. The Augustinian model I sketched above is one especially appealing version of such a restricted libertarianism, since it involves the transfer of responsibility via higher-order volitions, by which we may become responsible even for actions that we could not avoid causing.⁶⁴

V. A Rejoinder to Frankfurt

In his response to a much earlier and shorter conference version of this essay, Frankfurt asked why we should think that the conditions of responsibility for inner character should be any different than the conditions of responsibility for particular actions. Surely, he said, it is "far more reasonable to presume that the basic structural conditions for moral responsibility are the same in every type of case."⁶⁵ So if we can be responsible for acting on a given first-order desire even when we cannot do or desire otherwise, then likewise, we can probably be responsible for a volitional identification formed in the higher-order will even when we could not will otherwise. Frankfurt's suggestion is that there is a natural analogy between responsibility for these different elements of agency: if so, then such hierarchical or order-symmetry would be the natural presumption. Thus by contrast, hierarchical asymmetry in any theory of moral responsibility will be *ad hoc*, unless that theory can give us special new reasons to think (against our alleged natural presumption) that asymmetry is the best explanation of available evidence.

The libertarian rejoinder to this argument should now be evident. Even on Frankfurt's own view, the conditions of responsibility for particular outward (or non-reflexive) acts and responsibility for volitional identification are *not* structurally similar, because responsibility for outward actions traces (at least in part) to responsibility for volitional character and for the reflexive activities that constitute it. The hierarchical asymmetry in the Augustinian model builds on this crucial structural distinction: in this model, (1) responsibility for a particular outward act may exist even when the act is inevitable, but *only because* the responsibility flows from a higher-order volition through which the agent identifies with that act. Yet (2) responsibility for volitional identification itself does not trace to other psychic states or processes with their own distinct conditions of imputability, since this is where 'the buck stops.' This is just to say that the higher-order will (or some part of it) is a source of ultimate responsibility in Kane's sense. (3) Hence inevitability and responsibility cannot be made compatible for the higher-order will in the way they are rendered compatible at the level of outward-actions and their first-order motives. (4) Finally, note that we could add further tracing conditions at this level, if we think that sufficient persistence in our higher-order volitions or sufficient commitment of the will can 'fix' our identifications or make them impossible for us to change in the future.⁶ Of course, this does not show that compatibilists cannot offer competing accounts of responsibility for volitional identifications. But it does show that the order-asymmetry of the early Augustinian theory is hardly *ad hoc*, since it is just a specific instantiation of the structural asymmetry that we must find in *any tracing theory* (including compatibilist ones) between conditions of local responsibility for elements covered by tracing principles, and conditions of ultimate responsibility for those final elements to which all responsibility traces. Nor is the burden of proof on libertarians to demonstrate that the hierarchical asymmetry permitted by PAP-C is better than any conceivable compatibilist explanation of our experience and our moral concepts. Rather, we only have to show that hierarchical asymmetry has a clear and logical rationale in the elegant structure of an unrefuted theory with a respectable historical pedigree, which for most people will sufficiently explain all the relevant available evidence. We have made some progress in showing this much.

Pace Frankfurt, then, its hierarchical asymmetry is not *by itself* any *prima facie* evidence against the liberty of identification model. Hence the tracing libertarian can safely hold that, although examples like the willing addict show that responsibility may be locally compatible with inevitability for outward actions, this does not apply by analogy to the very volitional identifications that make such local compatibility possible. To think that we can simply generalize from such local compatibility is to miss the insight of the tracing theory that there is a crucial disanalogy between levels of the will: if a person is responsible for acting on inordinate desires even when they cannot do otherwise, but only because the person was committed to these first-order dispositions by her higher-order volitions, then this still leaves *an open question* about the conditions of responsibility for these volitions themselves. Plausible Frankfurt-style cases necessarily accommodate the fact that ultimate responsibility for the most central activities of human

agency can have conditions quite different from the conditions of derivative responsibility for lower elements of our agency.

Moreover, starting from the libertarian intuition embodied in PAP-C, we should expect to find another closely related asymmetry: even if the actual *causal history* of the first-order motive that explains an action may not always by itself determine the agent's responsibility for that action (e.g. if the agent identifies with a motive derived from external causes), yet the causal history of higher-order volitions may well be crucial in determining the conditions of responsibility for volitional identifications themselves. How such volitions are formed in the higher-order will may be essential to their special significance as carriers of the agent's identification. Since agents are by definition responsible for their volitional identifications, such a volitional attitude may not count as an identification (or as a self-determined expression of our long-term goals, ends, and values) unless it has the right kind of causal history, e.g. perhaps being brought about by a process of practical reasoning involving libertarian choice or agent-causation, or some other incompatibilist moment.

If the hierarchical tracing model with its order-asymmetry is established as an intelligible option that embodies PAP-C, we might ask why rival compatibilist accounts of responsibility for the higher-order will should still have any appeal, even if incompatibilists cannot find decisive defeaters for them? Since today the fear that future physics could describe our world as a deterministic system seems rather remote (especially given the successes of Superstring Theory) I think the compatibilist accounts will seem forced unless their defenders can actually refute PAP-C. For example, without such a refutation, why would we accept, as Frankfurt has asserted, that a person can be programmed to form the identifications he does, thus taking responsibility for various desires even though he could not have done otherwise than to identify with them or take responsibility for them?⁷ In his response to my paper, Frankfurt reiterated this thesis:

Whatever mental acts are required to accomplish the acceptance in which identification consists, a person may be morally responsible for performing those acts even if he could not have done otherwise than to perform them. There is no reason why it should be any more essential for an agent to have alternatives to *these* acts [which constitute an identification], in order to be morally responsible for them, than it is essential for morally responsible agents to have alternatives to acts of other sorts.⁸

But, given my above analysis, I do not see how Frankfurt could try to convince anybody to agree with this assertion if (like most people) they start with the presumption that PAP-C is correct. Again, I do not attempt here to demonstrate that Frankfurt's assertion is incorrect, but I have shown that he is wrong to think that his examples against familiar unrestricted PAP-principles legitimize any *presumption* that PAP-C is probably also false. Traditional (non-global) Frankfurt-style cases neither falsify PAP-C nor ground the least doubt of its truth. And since for most people, the presumption remains on the side of PAP-C unless it can be refuted, the

burden of argument is thus on Frankfurtians to provide quite different counterexamples against PAP-C itself.

In that effort they will face the following dilemma: (1) either their counterexample will depend on highly controversial inevitability-making conditions to ensure that the agent identifies as she does in the actual scenario, such as hard knowledge of what an agent's future decisions will be,⁶⁹ or (2) it will no longer preserve the difference between the *R* factors in the actual sequence that make the agent responsible for her identification, and the *I* factors that render it inevitable, in such a way that responsibility for *R* can be avoided. In the latter case, I have argued that libertarians need not grant the fundamental premise in traditional counterexamples to PAP, namely that the agent remains responsible for her activity.

A final example may help to clarify this problem for Frankfurtians. The film *Leaving Las Vegas* features Benjamin, a willing addict (played by Nicolas Cage), and while we pity him, most viewers probably blame him for identifying with his alcoholism.⁷⁰ But would we feel the same if we were told that Benjamin could not have done otherwise, because if he had not decided to quit his job and throw himself unreservedly into pursuit of death through drink, a counterfactual intervener would have *made* him form the higher-order volition expressed in these decisions? Call this case FV (for Frankfurtian-Vegas). Since the counterfactual intervener played no role in the actual sequence, should we say that in FV, Benjamin is responsible for his higher-order will, because he formed it on his own? Perhaps this will seem plausible until we ask: how does this differ from the intervener's *causing him* to form his volitional identification, as would happen if Benjamin failed to form it by a certain time, or began before then to turn towards becoming an unwilling addict?

As Fischer and others have repeatedly emphasized in rejoinder to flicker-of-freedom defenses of PAP, in traditional Frankfurt-style cases, there is an important asymmetry between the actual and counterfactual sequences: the agent is *not* responsible for what she does in the counterfactual sequence, because it does not amount to an action at all (instead, it is an unintentional movement).⁷¹ Even in cases where the counterfactual intervener supposedly causes a decision or an intention to act, the 'action' in the counterfactual sequence is still not voluntary, and the Frankfurtian cannot not hold the agent responsible for it without begging the question against incompatibilism. But now, in FV, we are supposed to imagine that the counterfactual intervener will cause a *state of volitional identification* to arise in our agent. According to the tracing theory, and Frankfurt himself, the agent is *eo ipso* responsible for any such state of his inner character: this must be the case for responsibility to transfer through this state to other first-order motives. So FV asks us to imagine that an intervener can cause the *very same state* for which the agent is responsible in the actual sequence. Thus in FV, the required asymmetry between actual and counterfactual sequences is lost: in both, Benjamin must be responsible for his higher-order volition. As a result, FV simply begs the question against the incompatibilist. So the Augustinian tracing-defender need not countenance that FV is even a possible scenario. Thus anyone beginning from PAP-C and then reflecting on these cases would be justified in concluding that in a scenario in which we can no longer

trace the compulsive agent's responsibility to a volitional commitment or identification that he could have voluntarily avoided, it would be unfair to consider him responsible for his fate.

If some non-philosopher NP (whose intuitions aren't shaped by explicit prior theoretical baggage) still did feel inclined to blame Benjamin for identifying with alcoholism, even though a counterfactual intervener was standing by to make Benjamin so identify, NP's attitude would probably be due to her tacit assumption that in the counterfactual sequence, the artificially created second-order desire to cultivate alcoholic cravings would not count as *the agent's own*, or as an identification for which he has taken responsibility. In other words, NP would be reinterpreting the FV case so that in the counterfactual sequence, Benjamin only acquires some *simulacrum* of volitional identification with his desire for alcohol. In this alternative version of FV, the intervener cannot really make Benjamin into a willing addict without his consent. In this form, FV would include the required asymmetry between actual and counterfactual sequences, and thus would not beg the question against incompatibilists. But in this form, FV also leaves PAP-C intact. For now Benjamin *can* voluntarily avoid identifying with his alcoholism by deciding on his own not to identify with it in the counterfactual sequence.

VI. Conclusion

So far, it seems that the tracing version of restricted libertarianism resists traditional Frankfurtian counterarguments. But two important questions remain to be addressed. First, do Augustine's mature or final views about freedom, Christian justification, and the human relationship with God undermine the sort of theory I've sketched here? In a sequel to this paper, I hope to defend the thesis that, while Augustine did change his views following the Pelagian controversy, his own earlier approach remains superior. Second, quite apart from theological concerns, can the proponent of restricted libertarianism extend the tracing defense of PAP-C to so-called 'global' Frankfurt-style cases, in which the agent never had the power to do otherwise at any point in the past, but still apparently acted without outside interference in the actual sequence? There are now several kinds of these cases, some of which involve complex problems. But the present paper has at least shown that to be *fully global*, such cases will have to include the higher-order will. Yet as we have seen, if the higher-order will is made inevitable by some series of interveners, then they must have the power to cause an agent to form a volitional state for which she is necessarily responsible, a state such that responsibility is essentially built into it. And it is hard to see how this could be stipulated without begging the main question that will divide libertarians and non-libertarians over PAP-C.

Fordham University

NOTES

1. Part of an earlier version of this paper was presented at the Eastern

Division meeting of the American Philosophical Foundation (Atlanta, GA: December 29, 1996). I wish to thank Professor Frankfurt for his helpful response, and for other reactions and advice in earlier correspondence.

2. Frankfurt, "Alternative Possibilities and Moral Responsibility," *Journal of Philosophy*, 66 no. 23 (December, 1969) and "Freedom of the Will and the Concept of a Person," *Journal of Philosophy*, 68 no. 1 (January, 1971); both reprinted in *The Importance of What We Care About* (Cambridge University Press, 1984).

3. Moreover, of the two philosophers who have connected Augustine and Frankfurt, Eleonore Stump and David Hunt both hold that moral responsibility only requires a form of incompatibilist freedom that does not require true libertarian (or alternate-possibilities) freedom. David Hunt bases his hyper-incompatibilist account of moral freedom on global overdetermination cases, some of which he derives from Augustine's own work on divine foreknowledge.

4. See Davenport, *Review of Responsibility and Control*, by John M. Fisher and Mark Ravizza, S.J. (Cambridge University Press, 1998), in *Faith and Philosophy* 17 no. 3 (July 2000): 384-395.

5. All references to Frankfurt's "Freedom of the Will" paper will be given parenthetically by page number in *The Importance of What We Care About*.

6. So understood, "alienation" is simply the negative correlate of identification. The unwilling addict alienates the first-order desire (a craving for drugs) on which he nevertheless acts. That craving still becomes his first-order will in Frankfurt's sense, meaning that it determines the intention through which we would explain the addict's behavior as a particular type of action (e.g. 'getting high' rather than 'taking medication for an illness'). This his 'getting high' constitutes an intentional action, guided by his own understanding of what he was doing: it is not involuntary, like sleepwalking. But nevertheless, the very intention that makes this an act of 'getting high' is in a deeper sense against his inner will, or alien to his self.

7. Augustine, *On Free Choice of the Will*, tr. Thomas Williams (Hackett, 1993), Book III, §1, p.72.

8. In the thirty years following Frankfurt's 1971 paper, many critiques have made it clear that his initial analysis of volitional identification in terms of a kind of second-order "desire" was too simple. Frankfurt originally conceived second-order volitions as simply one kind of iterated desire: a "desire to desire to X." But identification must consist in more than such an iterated preference or pro-attitude, because even complex preferences can be 'alienated' from the agent. Moreover, to avoid regress objections, the higher-order act that constitutes identification must be differently characterized to make clear why it too does not require endorsement or acceptance from some yet-higher order attitude to make it the agent's own. For details see my Ph.D. thesis, *Self and Will* (University of Notre Dame, 1998). But these concerns, which are central to the problem of explaining autonomy, are not crucial for the topics of this paper.

9. Augustine, *Free Choice of the Will*, Book I, §4, pp. 7-8. As a result, he says, "All wicked people, just like good people, desire to live without fear," but the wicked desire to possess securely or without fear goods that are essentially dependent on contingency, such as continued biological life: thus someone who murders another in fear of them "desires to live without fear" (p.6) but applies this desire to a good for which it is 'inordinate,' such as physical well-being. If this seems to imply that killing in self-defense would be murder, Augustine says that in such cases we can act for the sake of the principle (which is not subject to moral luck) that "it is much worse for someone unwillingly to suffer a sexual assault, than for the assailant to be killed by the one he was going to assault" (p.8). Acting on this motive would be inordinate.

10. The idea behind the concept of "inordinate desire" seems to be that anything whose possession and enjoyment is temporary or otherwise substantially dependent on luck can have intrinsic value, but only of a lower order than the value of the different virtues. Inordinate desire is a willingness to pursue these lower values at the cost of the higher values. It is unnatural to value such luck-dependent goods so much that we are willing to commit injustices in the pursuit of them. Still, the idea that luck-dependence by itself places something's value on a lower order than qualities of character that are not luck-dependent involves a Stoic attitude that not everyone will endorse.

11. Augustine, *Free Choice of the Will*, Book I, §11, p.17. Augustine arrives at this conclusion by an elimination argument. He has already argued that a divine power, which is superior to the changeable mind and will distinctive of persons, would not compel us to act on an evil motive, and that no power inferior in nature to our minds and will can make them "slave to inordinate desire" (p.17). Therefore by elimination, it can only be our own will that does it. Of course, given our contemporary knowledge of physical addictions, we might want to qualify Augustine's claim today.

12. *Ibid*, Book III, §17, p.104.

13. *Ibid*, pp.104-5.

14. *Ibid*, §25, p.121.

15. *Ibid*, §24, p.120.

16. *Ibid*, §25, p.122.

17. It seems that Carlos Steele misses this aspect of Augustine's account in his otherwise brilliant essay, "Does Evil Have a Cause? Augustine's Perplexity and Thomas's Answer," *Review of Metaphysics* 48 no.2 (December 1994): 251-73. But he is correct that Augustine's own example of stealing the pears "seems to undermine the fundamental premise of Thomas's philosophy of action, namely that nobody loves sin and evil for themselves" (p.268). At least, it seems that the human person is capable of valuing evil intrinsically as an expression of absolute denial of others, as a way of trying to reject our nature as beings already responsible for others.

18. St. Augustine, *On the Trinity*, Book X 11.18, reprinted in *The Essential Augustine*, 2nd Ed., ed. Vernon J. Bourke (Hackett Publishing, 1974): p.77 (my italics). In the next sentence, Augustine adds that "nothing is so much in the memory as memory itself"—which parallels his earlier claim about the reflexivity of will.

19. Augustine, *Free Choice of the Will*, Book I, §12, pp.19-20.

20. *Ibid*, Book I, §12, p.19.

21. In Kant, the only difference is that Augustine's eudaimonistic criterion for good first-order intentions and desires is replaced by formal universalizability tests for the moral acceptability of the "maxim of the act" or the first-order will involved in the action *qua* action.

22. Augustine, *Free Choice of the Will*, Book I, §14, pp.23-24.

23. Bonnie Kent traces the history of this fundamental Augustinian idea through medieval philosophy in the 13th century, culminating with Duns Scotus's distinction between the will to happiness and the will to righteousness. See Kent, *Virtues of the Will* (Catholic University of America Press, 1995).

24. *Ibid*, Book I, §11, p.18. In the dialogue, it seems incredible to Evodius (who later defends the Platonic intuition that no one wills to be unhappy) that anyone could have an evil will in this sense, but he acknowledges that the story of the Fall implies it. Augustine seems to hesitate, however, on whether anyone could will to be moved by inordinate desires just because of their inordinateness. This would be wickedness for its own sake. Yet his later account of the rebellious will in Book III suggests something close to this.

25. This does not mean, however, that actions as opposed to states of *will* in the Augustinian sense are mere behaviors that do not essentially involve a description of their intended purpose. Rather, *intentional* behavior or 'human action' in the Aristotelian sense is distinct both from mere behavior (which may not constitute an action), and from higher states of volition. The latter are states that aim at cultivating or extinguishing various kinds of first-order motives that explain the agent's intentional actions.

26. St. Augustine, *Confessions*, tr. R. S. Pine-Coffin (Penguin Books, 1961, 1986 pb.), Book VIII.5, pp. 164-6; my italics for emphasis.

27. *Ibid.*, Book VIII.8, p.171: "...for to make the journey, and to arrive safely, no more was required than an act of will. But it must be a resolute and wholehearted act of the will, not some lame wish which I kept turning over and over in my mind, so that it had to wrestle with itself..."

28. See Eleonore Stump, "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will," in John M. Fischer and Mark Ravizza, eds., *Perspectives on Moral Responsibility* (Cornell University Press, 1993): 211-236, esp. p.229. I am indebted to Professor Stump for advice on this paper and for helpful discussions of Frankfurt's hierarchical model, the doctrine of justification, and other questions. For more on weakness in the higher-order will, see my essay, "Kierkegaard, Anxiety, and the Will," *Kierkegaard Studies Yearbook*, Vol. 6, ed. Niels Jørgen Cappelørn, Hermann Deuser, and Jon Stewart (Berlin: Walter de Gruyter, fall 2001): 158-81.

29. Augustine, *Free Choice of the Will*, Bk. III, §1, p. 72.

30. Although the term "libertarianism" is now widely used as a synonym for "incompatibilism," referring to any understanding of the freedom required for responsibility that is incompatible with psychophysical determinism, I continue to use "libertarianism" exclusively for positions which are incompatibilist at least in part because of their alternative-possibilities requirements among the conditions of moral responsibility for actions, omissions, decisions, intentions, states of character, etc. "Libertarianism" is the best short label we have for such theories, so it is useful to reserve the term for them.

31. *Ibid.*, Book I, §13, p.20.

32. *Ibid.*, Book III, §1, p.71.

33. Kantians called this process a free act of *willkür*, meaning that it is causally undetermined. But a libertarian account of the higher-order will need not conceive the process of election as a spontaneous 'choice' in Kant's sense of a noumenal act.

34. For a detailed development of this point, see Davenport, "Towards an Existential Virtue Ethics," new in *Kierkegaard After MacIntyre*, ed. Davenport and Rudd (Open Court Publishing Company, 2001): 265-323, esp. §III.

35. Bourke, *The Essential Augustine*, p.176.

36. See Hunt's paper, "On Augustine's Way Out," *Faith and Philosophy* 16 no. 1 (January 1999): 3-26. I am especially indebted to David Hunt for pressing me to see that this question cannot be avoided if we are adequately to think through Augustine's relationship to Frankfurt. However, for reasons of space, my response has been postponed for the intended follow-up to the present paper.

37. There is an ambiguity in Frankfurt's concept of the wanton, since he does not distinguish between beings who have the *capacity* to form higher-order volitions but have not exercised this capacity ('voluntary wantons') and those who are *essentially* wanton (or wanton in every logically possible world in which they exist). In my example, we have a third kind of wanton: a being who is not essentially wanton, but who cannot voluntarily overcome his wantonness. For this kind of wanton, like essential wantons such as dogs and cats,

it seems that the categories of moral worth do not apply. Such wantons are *beneath* both good and evil in the will.

38. Of course, someone may object that the phrase "could have been otherwise" is sometimes used in other ways. But Frankfurt typically uses it to designate the power at a given time, in one's present condition, to bring about or to actualize alternate possibilities—for example, in referring to PAP as the principle that responsibility entails that the agent could have done otherwise than he did.

39. Formally, we may define Frankfurtian freedom of the will as (1) liberty of identification along with (2) the condition that actual second-order volitions control which first-order motive that agent acts on.

40. Augustine, *On Free Choice of the Will*, §16, p.16.

41. *Ibid.*, Book III, §19, p.107.

42. *Ibid.*, p.118. Moreover, the reason for our original capacity to sin or turn away from God (in addition to our natural capacity to love God) is the soul's self-consciousness, in which it "realizes that it is not the same as God" (p.122). Elsewhere, Augustine also suggests that it is because the human soul was "made from nothing" that it is capable of turning itself from God to something of lesser worth (e.g. itself, in pride). See, for example, the last paragraphs of *City of God*, Part Three, Book XII.

43. *Ibid.*, Book III, p.107.

44. *Ibid.*, Book III, §20: "Yet if anyone was willing to turn back to God so that he might overcome the penalty that had been imposed for turning away from God, it was right for God not to hinder him" (p.108).

45. This could work in the opposite way as well, as Kant emphasized: if someone acts on sympathetic feelings towards others but does not *actively identify* with this impulse or incorporate it because it motivates an objectively good action, only a hollow semblance of virtue is present. We might even imagine an actively *unwilling* sympathetic person—perhaps he has become convinced that his desire to help others is merely a symptom of 'co-dependence' that should be overcome, or that it causes him to make too many sacrifices.

46. Frankfurt, "Identification and Wholeheartedness," reprinted in *The Importance of What We Care About*, p.170.

47. In later papers such as "Identification and Wholeheartedness," Frankfurt does agree that the formation of second-order volitions must involve some *practical reason*: it cannot be totally arbitrary or indifferent to the person (pp.168-9). But his analysis of *caring*, which involves identification, implies that in some cases we may not be able to avoid forming the higher-order volitions we have.

48. In addition to the later papers in *The Importance of What We Care About*, there are several papers bearing on this theme in Frankfurt's new collection: *Necessity, Volition, and Love* (Cambridge University Press, 1999).

49. These cases were first introduced in Frankfurt's paper, "Alternate Possibilities and Moral Responsibility." They include, for instance, cases where a 'counterfactual intervener' wants the agent Jones to do a certain act A, and is prepared to make Jones do A if Jones shows any sign of doing B instead, but in the actual sequence Jones deliberates and chooses to do A on his own anyway, and so the intervener is not triggered. Such cases, along with many variants, have received extensive discussion in recent literature, without final resolution to this day.

50. Davenport, "Review of *Responsibility and Control*, by Fischer and Ravizza," p.386.

51. I have recently realized that a similar point is made by Laura Waddell Ekstrom in her article, "Protecting Incompatibilist Freedom," *American Philosophical Quarterly* 35 no. 3 (July 1998): 281-91, p.284. But Ekstrom does not

develop this point into a defense of a tracing version of PAP. Rather, she argues that if the actual sequence is deterministic, then we should not grant the initial assumption that agent A is responsible for X. Still, this coheres with the tracing defense's claim that for the agent to be responsible for conditions R, these cannot have been caused by deterministic laws and initial conditions preceding the agent's birth.

52. Susan Wolf has already anticipated the liberty of identification model in her book, *Freedom Within Reason* (Oxford University Press, 1990), ch.3. Arguments that libertarian freedom will involve arbitrary decisions or random moves require a different sort of response, which I hope to provide in future work. Robert Kane has offered a libertarian explanation of why, as Wolf says, "some decisions we make seem neither inevitable nor arbitrary" (p.48) and I think his account can be improved by arguing for a kind of synthesis of Wolf's Autonomy and Reason views.

53. John M. Fischer and Mark Ravizza, S.J., *Responsibility and Control* (Cambridge University Press, 1998), ch.8.

54. However, I think the source problem and weakness of will prove to be defeaters of compatibilist theories. See my review of *Responsibility and Control* and my paper, "Fischer and Ravizza on Moral Sanity and Weakness of Will," *Journal of Ethics*, 6 (2002): 235-59.

55. Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), p.35.

56. *Ibid.*, p.33.

57. *Ibid.*, p.35.

58. *Ibid.*, p.72.

59. In my view, this feature of Kane's account may make UR more vulnerable to the same kind of Frankfurt-style objections which have been raised against the key transfer principle in Peter van Inwagen's so-called "direct" argument for the incompatibility of moral responsibility and determinism. Thus the alternative theory I favor does not depend in this way on the idea of transfer of responsibility for X to all actualized sufficient conditions of X.

60. This is probably because Kane's theory emerges primarily from reflection on van Inwagen's Consequence argument, while my tracing defense strategy emerges directly from reflection on the Frankfurt-style cases.

61. Kane, p.142. Also see Kane, "Replies to Fischer and Haji," *Journal of Ethics* 4 (2000): 338-342; "Carl Ginet and I, as well as David Widerker and others, have argued that Frankfurt-controller scenarios, and hence Frankfurt-style examples, do not work in indeterministic worlds where choices might be undetermined" (p.341). This is a version of what I have called the trigger defense against F-style cases. In some more recent papers, however, Kane suggests something closer to the tracing defense for non-global F-style cases, but then he reverts to the trigger defense against global Frankfurt-controllers to argue that they cannot control will-setting acts or SFAs: e.g. Kane, "Freedom, Responsibility, and Will-Setting," *Philosophical Topics* 24 no. 2 (Fall 1996), *Free Will*, ed. Christopher Hill: 67-90, pp.85-86; and Kane, "The Dual Regress of Free Will and the Role of Alternative Possibilities," *Philosophical Perspectives* 14 (2000), *Action and Freedom*: 57-79, p.64, p.76.

62. Stump, "Alternative Possibilities and Responsibility: The Flicker of Freedom," *The Journal of Ethics* 3 (1999): 1-26, pp.11-12.

63. This is the route I explored in my paper, "Global Frankfurt-Style Cases: A Libertarian Reply," presented at the Pacific Division meeting of the *American Philosophical Association* (Seattle, WA, March 31, 2002).

64. Against Frankfurtian hierarchical compatibilism, Kane effectively presses the source-problem, or objection about manipulation of the higher-order

will by covert non-constraining controllers (*The Significance of Free Will*, pp.61-71). While I agree with this argument against Frankfurt, I've argued that if volitional identification is understood as involving libertarian choice, it can play a very useful role in a robust libertarian theory. Forming volitional identifications, or the cares and commitments of which they are parts, may be among the most important kinds of "will-setting" processes in Kane's sense.

65. Harry Frankfurt, "Reply to Davenport," Presented at the Eastern Division meeting of the American Philosophical Foundation (Atlanta, GA: December 29, 1996), p.1.

66. This extension of the Augustinian model may help address some of Frankfurt's claims in papers such as "The Importance of What We Care About" and "Rationality and the unthinkable" that some second-order volitions can be part of a particular agent's built-in nature, or can become so fixed that they are volitionally necessary for that agent, and thus unchangeable by him. Existentialist accounts of personhood reject the claim that persons have 'individual volitional natures' in just this sense, precisely because of their perception that liberty of identification is essential to personal existence as such.

67. Harry Frankfurt, "Three Concepts of Free Action," *Proceedings of the Aristotelian Society*, Supplementary Volume (1975); reprinted in *The Importance of What We Care About*: 47-57, p.53.

68. Frankfurt, "Reply to Davenport," p.3.

69. See, for example, the fantastic coincidence, divine omniscience, and Newcomb's predictor cases suggested by David Hunt in his article, "Moral Responsibility and Unavoidable Action," *Philosophical Studies* 97 (2000): 195-227, pp. 218-220. What is most significant about these cases for me is not that they dispense with a trigger, but that they suggest new ways in which the I and R factors could be kept distinct. Here I can only promise to address such global cases in later work.

70. For, as Frankfurt has rightly said, identification may consist more in an "acceptance" of some features of one's motivational psychology rather than in a moral endorsement or approval according to objective criteria: "a person may identify himself with aspects of his character of which he actually disapproves but which he nonetheless, in a spirit of resignation, recognizes as integral to his nature" (Reply to Davenport, p.3). I would alter this only by suggesting that identification must still always involve a kind of commitment in which the agent makes a project out of maintaining and strengthening the motives with which she identifies. Despair or resignation will not by themselves constitute identification without this sort of commitment or active investment of the self. In other words, the despair must be volitional.

71. See, for example, Fischer, "Recent Work on Moral Responsibility," *Ethics* 110 (October 1999): 93-139, pp. 110-11.