Moral Responsibility, The Reactive Attitudes, and The Significance of (Libertarian) Free Will

(To appear in *Libertarian Free Will*, edited by David Palmer (2014)).

Dana Kay Nelkin

I. Introduction

It is often tempting to take it as a given that the topic of free will is an important and central one in human life, and then go on to engage in the rigorous debate about what having free will requires and whether we have it. Robert Kane's *The Significance of Free Will* takes nothing for granted. Though the book is perhaps best known for its wonderfully detailed and subtle development of a libertarian theory of free will, the task of the central chapter of the book, as the title also makes clear, is to ask and answer "the Significance Question." The question turns out to be a closely related pair of questions:

> The Significance Question: Why do we, or should we, want to possess a free will that is incompatible with determinism? Is it a kind of freedom "worth wanting (to use Dennett's useful phrase) and, if so, why? (13)

Or, to put it in other words, why have so many people--both traditional philosophers and others--wanted a kind of agency or authorship of their actions that requires indeterminism, and why should we? (80). Now I take it that one question is really psychological or etiological (what is it about people that makes them want this form of agency?), and the other normative (what is valuable about libertarian freedom?).

Kane's answers to these questions have clearly contributed both to framing and further driving the debate about the significance of free will. In what follows, I will set out those

answers, which raise a number of independently interesting issues, and which include an analysis of the free will debate itself, commitments to meta-ethical principles in value theory, and substantive psychological explanations for why we care about what we do. As Kane notes at the outset, "the usual answer" is that libertarian free will is necessary for a number of other things that people want and that are worth wanting. Candidates include: "(1) genuine *creativity*; (2) *autonomy* (self-legislation) or *self-creation*; (3) true *desert* for one's achievements; (4) *moral responsibility* in an ultimate sense; (5) being suitable objects of *reactive attitudes* such as admiration, gratitude, resentment, and indignation; (6) *dignity* or *self-worth*; (7) a true sense of *individuality* or *uniqueness* as a person; (8) *life-hopes* requiring an *open future*; (9) genuine (freely given) *love* and *friendship* between persons (or in religious contexts, freely given love toward God); and (10) the ability to say in the fullest sense that one acts *of one's own free will*." (80). But this is just the first step, and Kane makes clear that ultimately the answer lies in focusing directly on the value of the indeterministic origination, within one's self, of one's actions.

In section II, I set out and assess Kane's approach to the Question of Significance. In section III, I offer a different approach, and explore the application of the competing approaches to moral responsibility and the reactive attitudes thought to presuppose it.

II. Answering the Significance Question From the Value of Origination

According to Kane, identifying the things worth wanting that are thought to presuppose libertarian free will is the first step in the debate over the Significance Question. But, as Kane is quick to point out, this answer has not been accepted by compatibilists in the debate. Compatibilists tend to agree that these are all things worth wanting, but they disagree that

libertarian free will is required to achieve them. Rather, compatibilists offer "interpretations" of each of the ten items that do not require libertarian free will in particular. To that, incompatibilists have replied in turn that the compatibilist interpretations are not "what we *really* want." According to Kane, this is the first stage in the dialectic about libertarian free will. For example, when it comes to (1) on the list--creativity--Kane quotes Karl Popper, writing that determinism "destroys" the idea of creativity. It reveals an illusion in the idea that one has created a lecture, since instead "there was no more in it, according to physical determinism, than that certain parts of my body put down black marks on white paper" (81).[i] Or consider (3) and (4), desert and moral responsibility. Kane has us imagine a medical researcher whose discovery of a potent new antibiotic drug marks a great achievement. When colleagues say that it is all a matter of luck, since she was just in the right place at the right time, she is indignant, claiming that she is responsible for some of those very circumstances that led her to discover the drug. "What this shows," writes Kane, "is that the woman's conception of true desert for her accomplishments involves the idea that she is, to some degree, ultimately responsible for them," where ultimate responsibility entails indeterminism (82).[ii] Now substitute for the scientific discovery a morally praiseworthy act like a man's saving a drowning child. "He feels demeaned by the charge that he is undeserving because he just responded out of instinct, and would insist that the act "was not entirely caused by circumstances that were not up to him and not ultimately his doing." (83) Similar points apply, argues Kane, to blameworthy acts, particularly when punishment is involved.

Compatibilists respond to each of these points by claiming that in perfectly ordinary senses of the words, determinism does not preclude creativity or novelty, desert for achievements, or blameworthiness and praiseworthiness. For example, Beethoven's Fifth

Symphony would still be a remarkable new contribution to human history, even if determinism were true, and Beethoven would be admirably creative for having produced it. Of course, there is another notion of creativity that would be precluded by determinism, but that requires defining such a notion precisely in opposition to "determinism," and is not the ordinary notion. Similarly, when it comes to moral responsibility, in ordinary life, the default is to hold people responsible unless there are some excusing or mitigating conditions, such as coercion, ignorance, delusion, or manipulation.[iii] No mention is made of determinism. And, writes Kane, if moral responsibility can be understood in these terms, so can desert for achievements and more (90). At this point, Kane's incompatibilist will reply that the compatibilist is correct that there are notions that are compatible with determinism, but that there are more "exalted" counterparts that require indeterminism and that we also want (91). Yet Kane takes it that the compatibilist will reply that this last claim is question-begging. If the more exalted notions are just defined as incompatibilist versions of the ordinary notions, then the claim appears to build in the conclusion. Hence, we have what looks like a kind of stalemate.

The resolution, as Kane sees it, is to "look more closely at the notions of ultimate responsibility and underived origination themselves. Why *are* they so important? What lies behind the widespread conviction…that ultimate responsibility and underived origination confer greater value on each of the goods (1) to (10)?" (91) Compatibilists are often entitled to their notions for everyday life and the law courts, but when it comes to our place in the world, we must become metaphysicians, and ask these questions about the importance of concepts that build in indeterminism. (In what follows, I will use "indeterministic origination" to refer to ultimate responsibility and underived origination, for simplicity and to highlight the required indeterminism.)

Importantly, Kane here reverses the direction with which the debate, as described earlier, began. Rather than asking about the nature of what is claimed to be at stake and then enquiring into whether each phenomenon, precisely identified and evaluated, really requires indeterminism, we are to instead begin with the value of indeterministic origination itself. Kane's approach is also distinctive in the particular way in which it answers the Significance Question. For, as we will see, Kane offers an explanation for why we want to be undetermined originators in developmental psychological terms, and at the same time he explicitly omits to endorse the claim that we are all *rationally* required to value such undetermined origination.[iv]

Kane offers two lines of reasoning to explain why we value indeterministic origination, the first with roots in a developmental story about human experience of the self, the second relating it to the notion of objective worth. Reflecting on the experience of human infants who take great pleasure and a keen interest in what they cause by acts of their wills, Kane observes that this experience is what helps give infants a sense of self, distinct from the rest of the world. As our understanding grows, we become aware that we are influenced by the world in myriad ways and that we are physical objects just like those in the rest of the world. This provokes a "spiritual crisis", according to Kane. And our worries about determinism can be seen as an extension of this sort of doubt. It is "*a stage of the dialectic of selfhood,*" a stage in a process in which we learn about our selves and our relation to the world, all the while trying to hold on to the idea that we are independent sources of motion in the world distinct from the forces of the external world. It is in this way, according to Kane, that the dialectic of selfhood leads us to the notion of indeterministic origination itself. Thus, at least one reason we want indeterministic origination is the same reason that children and adults revel in their accomplishments, even their

very earliest, namely, that they are thereby enabled to distinguish themselves from the world (98).

A second explanation for why we value indeterministic origination is its connection to the notion of objective worth. Because we value our own objective worth, we value such origination. Kane offers this example in illustration: Alan is a despondent artist. A rich friend who wants to lift his spirits arranges to have his paintings bought under an assumed name for a large sum of money. Alan believes that his paintings are being bought because of their artistic value and cheers up. He dies happy, but under an illusion. Kane then has us compare this world to one in which Alan really is a great artist.[v] Suppose that Alan could stand outside both worlds, would it make a difference to Alan which world he inhabits (not that he would know)? We can easily imagine that Alan would think his life more valuable if he inhabited the second, despite the fact that he would be "subjectively" happy in both. If we value objective worth, then we value indeterministic origination, because the latter requires us to view the universe from an objective perspective. If we care about our worthiness above and beyond our being pleased, then we want to be undetermined originators.

Thus far, we have two explanations for *why* we want undetermined origination. Do these explanations also justify the claim that such origination is objectively valuable? Do we have an answer to the second part of the Significance Question? Kane does not go so far. He claims that there cannot be an argument that we are rationally compelled to value such origination; rather, at this point, argument gives out and we can simply use these explanations to *commend* this kind of freedom (99).

While the explanations are intriguing, I believe the particular explanatory claims are stronger than the evidence warrants in the end. Thus, questions remain about just why, and even

whether, we value indeterministic origination.[vi] At the same time, as I will argue, the general approach to claims about objective value does not go far enough. While the restraint is admirable in itself, it is in tension with what I take to be intuitive judgments about other values--including many items on the list of what is supposed to be at stake in the debate.

To see why, consider the explanatory arguments first. The early developmental story about the will and our sense of self is both plausible and important. But the crucial question is whether we should see our concern with *determinism* as an extension of the desire to be willing actors in the world, a desire for something that underlies our very sense of ourselves as selves. This is the move that is not supported here. We can distinguish the things we do from the things that happen to us even if determinism is true. And we can establish our sense of a self, distinct from the rest of the world, by simply understanding that we control some things in the world. We do not in turn need to have control over the *source* of that control to distinguish ourselves and feel our effects in the world. It is a further philosophical question whether we need a sort of ultimate control for other things we might want--moral responsibility, the justification of punishment, the justification of the reactive attitudes, genuine friendship, and so on. But the account of the development of one's sense of self by itself does not support this strong conclusion.[vii]

The second line of reasoning takes the desire for indeterministic origination to be of a piece with the desire for objective worth. The particular argument contains a powerful intuition pump, and does pump (at least for me) the intuition that the value of a life depends on more than just a subject's feelings and attitudes. We should conclude both that there is a strong case for the idea that being under an illusion adds disvalue to a life, and that it is better to be a great artist than not to be.[viii] But neither of these points adds up to indeterministic origination--unless either

being undeceived or being a great artist is understood in a way that presupposes it. The first idea seems a non-starter, since it seems clearly possible to be undeceived with respect to one's (lack of) talent in a deterministic world. If we were to try to extend the second idea, it seems that the two worlds we should compare are these: a world in which Alan is a talented artist in a deterministic world and knows it and a world in which Alan is a talented artist in an indeterministic world and knows it. If he (and we) would choose the latter, this would show that we value indeterministic origination. But at least for me, this intuition pump does not have the force of the first, and, interestingly, it does not even appeal to objective worth *per se*.[ix]

We could instead compare two worlds that differ only according to Alan's *deservingness*, and perhaps this is what Kane has in mind. But here I think we could only reach the conclusion about our valuing of indeterministic origination of we already assume desert for artistic achievement requires indeterminism (as opposed to requiring only talent, hard work, dedication, and so on). This would be to put us back at what Kane sees as the first stage of the debate, rather than showing independently something more fundamental about the value of indeterministic origination.[x]

Thus, I do not think that either explanation is adequate to account for the desire for indeterministic origination. Importantly, though, attention to human development does contribute a great deal to our understanding of why *the will*--and control--are important to us. Nevertheless, once we leave behind the extended developmental account that includes attitudes toward *indeterminism*, it is less clear just how widespread the desire for indeterministic origination is. At the same time, Kane may give up too soon on the question of whether we can make more progress on the question of whether such origination is *worth* wanting. The question is important because the significance of the entire debate about free will depends on at least

something being worth wanting -- indeterministic origination or any of the ten (or more) things thought to depend on it.

At this point, it should be noted that whether it is rationally required for someone to actually value something is not the same question as whether it is actually worth wanting. (Thus, Kane's pessimism about the possibility of rational persuasion need not be the same as pessimism about objective value.) But the two questions are connected--if we are to show that something is worth wanting, we need to give reasons of some sort for it. If indeterministic origination turned out to be like a case in which someone enjoys collecting small bits of paper, and we cannot say why this activity is valuable beyond his liking it, then there would be a serious gap in our account of the significance of such origination, and it seems we would have left the question of whether it is worth wanting behind altogether.[xi] But there seems to be something in between being a mere matter of taste (like collecting paper, perhaps) and something that everyone is rationally required to value for himself or herself, and that might be captured by the idea that it is *reasonable* for some people to value it.[xii] For example, I may not value *my* being able to sing well in front of large numbers of people, but I understand that doing so is valuable and contributes significantly to lives worth living. What can it mean to be reasonable that goes beyond a matter of mere taste? One possibility is that it is one way among many to achieve something that is worth wanting for all--e.g., a life in which one employs one's talents to their fullest.[xiii] This seems to explain well our judgments about a number of lives that are intuitively flourishing ones, but quite different in their details. But it is less clear how to apply this idea to indeterministic origination. What would be the larger category of good for which this is one among other ways to achieve it? Without an answer to this question, we seem unable to go beyond the claim that our desire for indeterministic origination is a matter of taste.

It is a notoriously difficult question how we can show anything to have objective value, or to be rational, or even reasonable (in the sense just described) to value. But Kane's discussion shows us in a vivid way that we must confront such questions if we are to even to interpret correctly the second part of the Significance Question, or to see it as a well-formed one. While it may be very hard to elicit direct intuitions about the value of being indeterministic originators in the absence of other things which may depend on our so being, this very fact seems to me to point to the fruitfulness of the first general direction of debate--eliciting such intuitions about the value of other things and then figuring out whether they require indeterministic origination after all.[xiv] Ideally, we will be able to do more than elicit intuitions. While it may be that at some point, argument runs out in the sense that there is no *more* fundamental value on which the value of, say, friendship or love rests, that need not mean that argument runs out in the sense that we can't make further appeal to the value's explanatory power, connection to other commitments we have, and intuitive appeal.

In the second part of the paper, I compare the approach just assessed to the one that proceeds by first asking about the nature and value of those things thought to be at stake in the debate by applying both approaches to moral responsibility and the reactive attitudes that are thought in turn to depend on it (call this the "what is at stake" approach). I do not claim that the latter approach has (yet) given us a definitive answer, but I hope to show both some of the fruitful ways in which this alternative approach has already been pursued, and a potential direction for further progress. Finally, I return to an insight of Kane's, showing how it might be combined with the "what is at stake" approach in a promising way.

III. A Contrasting Approach and a Case Study

My approach contrasts with Kane's by beginning with what is at stake, asking about *its* value, and asking whether what is valuable really requires libertarian freedom (or even freedom at all). To illustrate the application of the two approaches, turn first to a central item on the list, possibly the one that has been most discussed in recent decades: moral responsibility.

According to Kane's approach, there are *two* notions of moral responsibility, one that requires indeterminism and one that does not. The compatibilist version works for many everyday purposes, but it is reasonable to want the "more exalted" variety. We cannot claim that everyone should value this, on pain of irrationality, but there is a deep developmental explanation for our desire, and it is connected with something that seems of obvious value--our sense of selves and understanding of the very distinction between the self and world. But if this explanation fails for the reasons described earlier, then not only do we not have an argument for why we should value it, we lack even a motivation for locating this phenomenon in the world.

On the "what is at stake" approach, we should begin with the nature and value of moral responsibility, and, separately, investigate whether it requires or presupposes indeterministic origination. Even here, there are a variety of ways we might proceed. For moral responsibility might be valuable in itself, or it might be valuable because of what it is in turn required for, or, as I believe, both. For present purposes, I will follow Kane--and many others--in accepting that moral responsibility is presupposed by the reactive attitudes, those attitudes of people that, in the words of Peter Strawson "are directly involved in transactions with each other; … of offended parties and beneficiaries…such things as gratitude, resentment, forgiveness, love and hurt feelings."[xv]

On some views, responsibility is simply constituted by the appropriateness of the attitudes[xvi], but I will make a more minimal assumption here. The appropriateness of the reactive attitudes depends on our being responsible agents. This assumption is consistent with responsible agency being constituted by wholly response-independent conditions. What is essential here is that the attitudes are appropriate only if their targets are morally responsible agents. If this is correct, then it follows that if the reactive attitudes and our susceptibility to them are themselves valuable, then the value of our being morally responsible is (at least in part) derived from the value of appropriate reactive attitudes. Thus, in what follows, I will approach the question of the value of moral responsibility itself through the question of the value of the reactive attitudes and our disposition to them.[xvii]

Even having narrowed our focus in this way, there remain multiple routes to travel. For example, the reactive attitudes might be valuable because in turn they contribute to still other things that are valuable, or they might be valuable in themselves. If the former, then they might contribute to other goods necessarily, or only contingently. The current debate so far seems to be centering on the idea that reactive attitudes are valuable for their contribution to further goods. And here there is great variety in the further goods identified; they include good social structures and features like decreases in violence and disharmony, as well as the very existence of personal relationships. There is seemingly vast consensus that these are things worth wanting, so the debate has largely focused on whether the reactive attitudes do in fact contribute to these goods. More precisely, the debate has largely centered on whether the reactive attitudes are the best--or only--way to achieve them.

Here I enter the debate in one of its aspects, focusing on the claim that there is an unavoidable (or at least for most of us, unavoidable) connection between the reactive attitudes

and personal, caring, relationships. For if it can be shown that the reactive attitudes are valuable because they (at least for most) make personal and caring relationships possible, then we would be able to show not only the value of the reactive attitudes, but also--in one way--of morally responsible agency itself.

At the same time that Strawson introduced the term, "reactive attitudes," he argued for its importance to the debate about free will precisely by claiming their essential connection to personal relationships (pp. 61-62). Without these attitudes that are reactions to the good or ill will of others and through which we demand good will from those with whom we stand in a variety of personal relationships, we are left, on Strawson's view, to adopt the "objective" attitude, to "see him, perhaps as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided" (p. 66). If these are our only alternatives, then the high value of the reactive attitudes does not seem to require much in the way of argument.

But that the objective attitude and the reactive attitudes that presuppose responsibility are exhaustive alternatives is an assumption that has not held up well. In fact, Derk Pereboom has offered a rich and subtle account of a whole set of other options, or what he calls "analogues" of the reactive attitudes.[xviii] Even if we were to set aside the reactive attitudes that presuppose responsibility, many other personal attitudes remain, including moral sadness, disappointment, resolve, and analogues to forgiveness and gratitude that retain many positive features of each while eliminating any presupposition of responsibility.

But this is not the end of the debate. For at this point, there remains a question of whether we *can* eliminate our susceptibility to, or even regularly disavow, the reactive attitudes

and still maintain our personal relationships. If we cannot--even if only because of our human psychological limitations--this suggests that such susceptibility has significant value, contingent though it may be. And, even if we settle this (at least partly) empirical question, there is a further question: are the admittedly personal relationships that would remain without the specifically reactive attitudes *as* valuable as relationships in which we are susceptible to such attitudes? I address each of these questions in turn.

Taking up Strawson's cause, Seth Shabo has recently addressed the first question with a qualified "no" (2012). While there may be exceptional people who can disavow the reactive attitudes and still maintain caring personal relationships, most of us cannot do so. Shabo argues that genuine loving relationships, such as romantic ones, are ones in which the participants *care* not only for the other person, but also care about the other's attitudes and feelings about them. That is, if one is not susceptible to hurt feelings if, say, one's partner is unfaithful, or shows one disrespect or unkindness, then this is a sure sign that one does not have the kind of caring relationship that is characteristic of romantic love. One has to care about the other's feelings and attitudes displayed in his or her behavior, and to do this one must be at least susceptible to hurt feelings. As Shabo puts it, one must be susceptible to "taking things personally," and having a certain degree of "emotional vulnerability" (p. 112). In turn, hurt feelings "often beget resentment" (p. 114). While the first connection--between a caring relationship and taking things personally--is claimed to be conceptual, the second--between taking things personally and the reactive attitudes--is a psychological generalization. An example is meant to support this:

> Learning from a trusted friend that her husband openly scoffs at her artistic ambitions behind her back, Caitlin's feelings are badly hurt…As her initial embarrassment subsides,

> the first stirrings of resentment set in. He has been so supportive of their friends' creative endeavors, why not hers? When she thinks that this is her *husband*, her resentment shades into moral indignation…(115)

We are then asked to imagine that Caitlin wants to disavow such feelings or resentment and indignation because she believes them to be morally objectionable. But it seems that she cannot do so without "emotionally divesting" in a way that not only sets aside the feelings of resentment and indignation, but also the vulnerability to the hurt feelings. (115)

As part of his response, Pereboom offers a kind of real case in which many are able to both take things personally and disavow and even eliminate reactive attitudes: that of parents of teenagers who in a normal developmental stage, show disrespect. As he writes,

> Very commonly, teenagers go through a period when they have attitudes of disregard and disrespect for parents, expression of which can result in deeply hurt feelings. But often such expressions of disregard and disrespect do not occasion the parent's resentment, but instead their disappointment and sadness. Although these emotions are not reactive attitudes, they are nevertheless manifestations of vulnerability on the part of the parent. Crucially, they are also personal, since the teenager's attitudes toward his parents matter to them in their own right, apart from the consequences of these attitudes for their interests. Often parents in such situations are also resentful, but frequently they are not. (pp. 15-16 ms)

Both kinds of cases--that of Caitlin and that of at least some parents of teenagers--are powerful, but each has limitations. As mentioned, even Shabo does not claim that eliminating or disavowing resentment in a personal relationship is impossible; only that it is difficult or impossible for most of us. But there is another limitation, as well. The very case that Shabo uses to make the psychological generalization plausible has a particular feature that is not present in a number of cases of hurt feelings, a feature that may be quite relevant to determining how far our reaction to it should generalize. The case is one in which Caitlin is wronged by her husband, and one in which we naturally think resentment warranted. It is one in which we ourselves may be indignant on her behalf--and this has nothing to do with *our* susceptibility to hurt feelings in the case; we don't even know Caitlin's husband. If we are trying to show that the special susceptibility to hurt feelings that characterizes personal relationships gives rise to resentment (for most of us), then it is worth seeking cases in which third parties do not have such feelings of indignation. Take a case in which one's friend postpones a dinner in order to help another friend. A third party might feel sorry for one, but not indignant. Still, one might be hurt, even as one recognizes that one's friend did nothing wrong and fails to feel resentment. Or take cases of jealousy, in which the object of a person's love does nothing wrong. Or cases in which a child does not want to spend time with another child. These seem to be common sorts of cases in which one can have genuinely hurt feelings and yet not have the reactive attitudes. Still, it might be argued that one must be able to take all sorts of things personally, notably including cases in which one is wronged, to participate in a significant personal relationship, and in *these* cases in which one is wronged, it is impossible for most people to detach resentment from taking things personally. Note that this would be a substantially modified argument, but let us accept for the moment the premise that one must be susceptible to hurt feelings in all areas, including wrongs.

At this point, we might return to the case of parents of teenagers, which would seem to be a kind of counterexample to the psychological generalization, now restricted to the cases in which we take *wrongs* personally, that we cannot (or almost never can) separate out resentment. Perhaps this kind of case shows that in a large number of real cases of genuine wrongs, people are perfectly well able to separate out resentment from taking things personally.

But even if we are as optimistic as Pereboom about how commonly parents of teenagers are able to react without resentment and instead with disappointment and sadness, we can ask--as Pereboom does--whether this special ability is unique to the parental relationship. It also may be special to the relationship between parents and *not fully mature* adult children, in particular, insofar as there seems to be growing recognition of at least partial excuse of the behavior of adolescents.[xix] There may even be some reason to view these actions as not wrong, or not as violations of obligations for the very reason that teenage behavior is seen as a normal and healthy developmental stage. Thus, though the case provides some evidence that we are psychologically capable of setting aside resentment in personal relationships, how far *it* generalizes is also an open question.[xx]

As Pereboom points out, there is reason to think that our psychological flexibility is greater than Strawson and Shabo believe because in many cases we have been able to disavow and even eliminate our reactive attitudes to criminals, the insane, and children, among others (ms 16). In addition, it simply seems to remain an open empirical question to what extent we could either eliminate or disavow the reactive attitudes, particularly in non-parental personal relationships. But this does not mean progress has not been made; to the contrary. We have been able to recognize the importance of a specific empirical question, and a substantial start to

the process of answering it. A more settled answer will contribute directly to the question of the value of free agency.

Further, suppose, for the moment, that we had settled this question so that we found that we could make substantial changes in the way we conduct our personal relationships, so as to set aside the reactive attitudes while maintaining the personal aspects of them. At this point, we would be left without an affirmative answer to the question of whether the reactive attitudes are valuable. We would know that they are *not* valuable insofar as they are required for maintaining personal relationships. But this does not show them *not* to be valuable. Of course, they might be valuable for some of the other reasons mentioned earlier; for example, because they contribute to social harmony and a reduction in violence.[xxi] But we might be able to make further progress on the question of value even while continuing to consider personal relationships. To do this, turn to the second question set out earlier--assuming that we could maintain personal relationships in which we care and are disposed to take things personally, would we thereby lose something else of value? In particular, would we lose something of value in itself, namely, a particular *kind* of personal relationship? The question is whether relationships in which we are susceptible to the reactive attitudes have a special value that isn't equaled by the value of other sorts of personal ones.

Here is one way to argue for an affirmative answer. The reactive attitudes are our way of making moral (and perhaps other) demands on ourselves and each other. Through the reactive attitudes we hold each other to our obligations to treat each other in certain ways. Some of these obligations seem to be partly defining of particular relationships--e.g., friends have special obligations to friends; parents have special obligations to their children. (This is, of course, consistent with our all having obligations to all fellow humans, and even to sentient creatures.)

The reactive attitudes embody the demand that others (as well as ourselves) fulfill these obligations, or the holding of others to the standards defining of the relationship.[xxii] Relationships in which we cannot appropriately hold others to standards (but can only encourage or express disappointment about their failures to meet them) and in which we do not live up to or fall short of meeting demands might be thought to be missing something of importance. It is a truism that it is a gift and a sign of respect to hold others to standards, and it is thought to be an admirable trait when we do so for ourselves. Not all truisms are true, of course, but there is something appealing in the idea that there is special value in relationships in which we do so. To elaborate, it is helpful to note that the idea of "mutual demand" is here to be contrasted with the idea of "expectations" in a sense of "prediction." Two friends might in fact let each other down a fair portion of the time, but it might be argued that part of what makes them friends rather than two people who happen to care about each other is that they have expectations of each other in a sense distinct from that of prediction. If this is correct, then friendship itself would seem to require the holding to standards or the making of demands. But even if we are more expansive in our understanding of friendship, so that even two people who care about each other, without holding each other to the standards of friendship, still count as friends, the very recognition of and holding to standards seems to mark a real distinction between two kinds of relationships. Thus, a weaker conclusion is not that friendship itself requires the holding to mutual standards, but that there is a kind of friendship that requires it, and it has a special value. [xxiii]

It is important to emphasize that the value of being disposed to the reactive attitudes is consistent with it being the case that relationships that contain less *actual* resentment and indignation are better than those with more. And it is consistent with it being the case that relationships that contain more positive and less negative ones are better. The idea is simply that

the disposition and recognition of a disposition to such attitudes marks a particularly valuable kind of relationship.

A full evaluation of the approach sketched here would require examining more closely each of the two main premises in the larger reasoning: that the reactive attitudes are the only, or best, way to hold others to demands; and that holding others or oneself to demands is of any special inherent value. Either might be questioned. Interestingly, though, in the context of the larger project of exploring the value of free agency, it would not be enough for an opponent to simply undermine the first premise. For if the holding to demands itself presupposes free agency, then even if we could show the particular reactive attitudes unnecessary as a vehicle for doing so, the larger point that *something* of significant value presupposes free agency, of which the reactive attitudes might be one vehicle, remains.

Here I believe we can see at work a key insight of Kane's--there are different *kinds* of relationships, and some may have a kind of value that at least in one aspect goes beyond that of others. Interestingly, the approach here diverges on two dimensions from Kane's. First, Kane is primarily concerned to argue that freely chosen love, say, is superior, or valued more by many (p. 88).[xxiv] In contrast, the suggestion here is that relationships that presuppose a kind of free agency via the disposition to the reactive attitudes that partly define the relationships. This view is perfectly consistent with the idea that the feelings and commitments of friendship and loving relationships are not *themselves* undertaken freely.[xxv] Second, where Kane describes the increase in value in personal relationships as depending on the indeterministic origination, this approach is neutral as to what the relevant free agency requires in the way of determinism or indeterminism. But the insight that there may be special forms of some of the items on the list that depend on a substantial notion of *free agency* is shared between these approaches.

In sum, we have seen two ways of arguing that the reactive attitudes, and our disposition to them is valuable, as well as questions that have been (and might be) raised about both. If either argument is ultimately successful, we will have shown not only the value of the reactive attitudes and moral responsibility, but also, through them, the dependence of certain kinds of valuable relationships on a sort of free agency. At the same time, although we won't have shown that the free agency in question must be libertarian, we will have shown that several interrelated and important things are indeed at stake in the question. Thus, we will have shown the significance of a robust notion of free agency, and the significance of the *question* of whether that agency requires indeterminism.

IV. Conclusion

We have seen two quite different approaches. Kane's approach in one way goes further: by concluding that there is special significance to *libertarian* free will. But that significance is limited in another way: the significance depends on the reasons why people want it, but not reasons it is *worth* wanting. The alternative is to start with things that seem to be "worth wanting," even of the "exalted" kind that require free agency, explore whether they really are worth wanting (or worth wanting more than possible alternatives) and then see whether they require libertarian free agency in particular. I have sketched only one part of the second, "what is at stake" approach here, namely, the exploration of the value of the reactive attitudes that presuppose moral responsibility and their connection to certain kinds of personal relationships. But I believe that it is enough to reveal that, despite a lack of consensus, we are far from a stalemate about questions of value. And as Kane's own work on the nature of free agency

perfectly exemplifies, despite a lack of consensus on this point, as well, we are engaged in a lively and fruitful debate.

References

Brink, David (2004) “Immaturity, Normative Competence, and Juvenile Transfer: How (Not) to Punish Minors for Major Crimes,” *Texas Law Review* 82, 1555-85.

Fischer, John Martin (2002/2003) “Frankfurt-Style Compatbilism,” in *Contours of Agency: Essays on Themes from Harry Frankfurt* edited by Sarah Buss and Lee Overton (MIT Press); reprinted in Watson (2003).

Frankfurt, Harry (1971) “Freedom of the Will and the Concept of a Person” reprinted in Watson (2003).

Hurka, Thomas (1993) *Perfectionism* (New York: Oxford University Press).

Jeske, Diane (2008) "Special Obligations", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/special-obligations/>.

Kane, Robert (1994) *Through the Moral Maze* (Armonk: M.E. Sharpe Publishers).

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

Kohn, Alfie (2005) *Unconditional Parenting*: *Moving from Rewards and Punishments to Love and Reason.* New York: Atria Books.

Mele, Al (2006) *Free Will and Luck* (Oxford: Oxford University Press).

Milam, Per (in preparation) *The Value of Free Will.*

Nelkin, Dana Kay (2011) *Making Sense of Freedom and Responsibility* (Oxford: Oxford University Press).

Nelkin, Norton (1995) *Consciousness and the Origins of Thought* (Cambridge: Cambridge University Press).

Nichols, Shaun (2007) "After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes," *Philosophical Perspectives* 21, pp. 405-428.

Nozick, Robert (1974) *Anarchy, State and Utopia* (Oxford: Basil Blackwell).

Pereboom, Derk (2001) *Living Without Free Will* (Cambridge: Cambridge University Press).

Pereboom, Derk (2009) "Free Will, Love and Anger," *Ideas y Valores* 59.

Pereboom, Derk (forthcoming) *Free Will Skepticism, Agency, and Meaning in Life*.

Popper (1972) "Of Clouds and Clocks," in *Objective Knowledge* (Oxford: Oxford University Press)

Shabo, Seth (2011) "Where Love and Resentment Meet," *The Philosophical Review* 121, pp. 95-124.

Strawson, Peter (1962/2003) "Freedom and Resentment," reprinted in Watson (2003).

Wallace, Jay (1994) *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press).

Watson, Gary (1987/2004) "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. Reprinted in Watson (2004b).

Watson, Gary (2004a) "Two Faces of Responsibility," in Watson (2004b)

Watson, Gary (2004b) *Agency and Answerability* (Oxford: Oxford University Press).

Watson, Gary (editor) (2003) *Free Will, Oxford Readings in Philosophy, Second Edition* (Oxford: Oxford University Press).

Wolf, Susan (1981) "The Importance of Free Will," *Mind* 90, pp. 366-78.

[i] Popper (1972), p. 222.

[ii] Ultimate responsibility, as Kane sets it out, is as follows: An agent is *ultimately responsible* for some (event or state) E's occurring only if (R) the agent is personally responsible for E's occurring in a sense which entails that something the agent voluntarily (or willingly) did or omitted, and for which the agent could have voluntarily done otherwise, either was, or causally contributed to, E's occurrence, and made a difference to whether or not E occurred; and (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y. (72)

[iii] A compatibilist might also add a counterpart for excuses when it comes to praiseworthy action. If the saving of the child Kane describes were a result *only* of instinct, so that it wasn't done even with a view to preventing a harm to the child, then compatibilists could also have a reason for withholding moral praise. But if the person saved the child because she saw immediately the point of doing so without having to think about it, then the intuition that this description of her action would be demeaning in the least, or detract from her praiseworthiness in a sense worth wanting is far less powerful.

[iv] Another perhaps surprising element is the concession that compatibilist notions are often all that is needed in ordinary life and in the courts. This provides some mutual support for the general approach. If one is willing to concede that ordinary notions of moral responsibility, say, don't implicate indeterminism, then the direct question about the value of indeterministic origination takes on a kind of priority. However, it seems that it is precisely our ordinary concepts and practices that are central in the debate. As Kane recognizes, too, for example, our very justification for punishment often seems to hinge on resolving the metaphysical question at issue one way or the other.

[v] As Kane notes, this example is similar to Nozick's "experience machine."

[vi] It is important to note that Kane recognizes that not everyone does so value indeterministic origination, referring to certain Buddhists as one example.

[vii] See Norton Nelkin (1995) for a detailed defense of this point. It includes a developmental story that appeals to key findings in the developmental literature, shares key aspects of Kane's story of the fundamental importance of the will to the sense of self (and the world), and yet explicitly rejects the idea that this story is continuous with our concern about determinism. (See pp. 299-315.)

[viii] At the same time, a different pair of worlds is needed to provide full support for the conclusion that living without illusion is valuable in itself. For the worlds described here differ not only in there being deception in one and not the other, but in the fact that Alan is a great artist in one and not in the other. To isolate the deception, considering a world like that of the film, *The Truman Show*, in which although inhabiting a world in which one has the love of his family

and friends is better for Truman than one in which he doesn't, it is still better overall to not be deceived. Interestingly, Kane suggests that in a third world in which Alan finds out the sad truth is worse than all worlds, but I do not have this a strong intuition about this case, at least not with respect to the case as described thus far.

[ix] See Fischer (2002/2003) who also rejects Kane's thought experiment, but for somewhat different reasons. He attributes to Kane the premise that compatibilist views are committed only to subjective value because he assumes compatibilism does not have a "historical" component, and is only concerned with the structural arrangements of mental states. Because, as Fischer points out, his own account includes a requirement that the responsible agent have taken responsibility for the mechanism that generates his action and because it requires that the agent have the right connections to the world, his view is not targeted by this argument (p. 205). I agree, and simply emphasize that even a non-historical compatibilist account (e.g., a reasons-responsive account without an additional historical condition) can accommodate the intuition that is pumped by the thought experiment.

[x] See Kane (1995) for the case of Alan as an intuition pump for our view of objection worth, but with no conclusion drawn about indeterministic origination.

[xi] It is true that desire satisfaction itself is a candidate for making the desired object worth wanting. And it seems right, as Frankfurt (1971) points out, that a desideratum in any account of free will should explain why we would want it. But if this were all we could say, then the significance of indeterministic origination, or free will, or anything on the list of what is at stake, would seem to be of a shallow sort.

[xii] See Mele (2006), which presents a view that seems to seek a middle ground, and also seems in the spirit of Kane's position. In particular, Mele considers what he calls, "Daring Soft Libertarianism," or DSL, according to which libertarian freedom is of a kind that some find more valuable than a compatibilist kind. This view shares with Kane's the rejection of the claim that everyone ought to value being undetermined sources of their actions. As Mele writes, "The soft libertarian need not claim that everyone--or every reasonable person--values being an indeterministic initiator." (98) But, just as a gambler might show a risk-averse person what the former finds appealing in gambling, a soft libertarian might do something similar for one who does not herself value libertarian freedom. The particular analogy used suggests that DSL takes libertarian freedom to occupy a place closer on the spectrum to being a matter of taste than to being something of objective value. Like Kane's idea of being able to "commend" indeterministic freedom, Mele offers us the idea that a gambler might be able to explain to others who don't share his desire why it is that he likes it.

[xiii] Kane seems to adopt something not unlike this view in (1994), where he takes it that a number of different lives are valuable, but no one is called for by all. The view can be described as a pluralist conception of an "objective list" view of well-being. See, for example, Hurka (1993).

[xiv] It is worth noting two additional reasons why one might continue with the original direction of enquiry in the debate. At the outset, it is possible to be more optimistic than Kane seems to be about the possibility of progress if we begin with the value of the items thought to be at stake. For example, for at least some items on the list--such as the reactive attitudes--there is arguably a univocal notion shared by compatibilists and incompatibilists alike, and then it seems natural to focus on the questions of whether that agreed-upon phenomenon does or does not require indeterminism, and whether the items really are worth wanting, once we have a better idea of their nature. A second reason to resist reversing direction targets the idea that the presence of

indeterministic origination is what is supposed to explain the exalted status of the indeterministic version of each item on the list. But one might ask why, even if we could show why indeterministic origination is desirable, we should thereby favor an indeterministic version of *friendship*, say, rather than simply a compatibilist version of friendship, together with the additional, but distinct, fact of our being undetermined originators. Still, neither of these points detracts from the importance of asking directly about the value of indeterministic origination, for even if there were no obvious connection to any items on the list, it would still be a question worth asking.

[xv] Strawson (1962/2004, p. 62). The connection between reactive attitudes and moral responsibility is often taken to mark the fact that the notion of responsibility in question is that of "accountability," a notion distinguished from "attributability." (See Watson (2001) for the introduction of these terms.) One might blame someone in the sense of attributing a moral fault to an agent. One might call a piece of behavior (e.g., a colleague's failure to return a book) "shoddy," for example. So far, this is only to blame in the "attributability" sense. But this is not yet to blame in a sense that presupposes that the agent deserves to be held responsible, that demands can appropriately be made, and so on. The reactive attitudes are thought to embody the "making of demands." (See Strawson 1962/2003 and Watson 2004a). Interestingly for our purposes here, these two notions are sometimes taken to correspond to compatibilist and incompatibilist notions, with accountability requiring indeterminism. But there are compatibilist and incompatibilist theories of accountability, so for now, I will treat the notion in a neutral way.
[xvi] See, for example, Wallace (1994). This is also a natural interpretation of Strawson, as well. (See, for example, Watson (1987/2004b)).
[xvii] In fact, I believe that moral responsibility is also valuable in virtue of what moral responsibility consists in. It is, of course, a matter of great controversy under what conditions we are responsible. I have elsewhere (2011) defended the view that we are responsible insofar as we act with the ability to do the right thing (or a right thing) for good reasons. Although it is controversial whether this captures the conditions of responsibility, it seems to me much less controversial that acting with this ability is valuable. A third way in which responsibility (whether moral or not) may be thought to be valuable is its being required for our living a life without illusion. If, for example, we must take ourselves to be free and responsible agents, then our failure to be such agents would condemn us to a life of illusion. I discuss this proposal in (2011). See Wolf (1981) for discussion about how this point is connected to the value of the attitudes themselves, and Milam (in preparation) for a valuable response.
[xviii] See, for example (2001), (2007), and (2009) and (forthcoming).
[xix] See Brink (2004), for example.
[xx] It is an interesting empirical question just how the process works for many parents of teenagers who behave in disrespectful ways. If (as seems sometimes to be the case) the recommended thought processes include thinking of the behavior as normal and healthy and ultimately in service of establishing adolescent's own identity, and so on, the fact that it is disrespectful becomes less salient. *These* sorts of thought processes would seem to have some effect not only in reducing resentment, but also in taking the particular behavior personally (since the behavior is, in a sense, "not about you" the parent). Still, this is just one route to the reduction of resentment, and there may be others that would leave one's "taking it personally" fully intact. Further, even if parents of teens with reduced resentment also take the relevant behavior "less

personally," it is a plausible hypothesis that this need entail nothing about a lower *quality* of the personal relationship. To the contrary, it is often recommended to parents that they take this approach. See, for example, Alfie Kohn (2005).

[xxi] See, for example, Shaun Nichols' (2007) argument for this empirical claim, and responses by Pereboom in (2009) and (forthcoming). This can be seen as a kind of parallel to the debate described in the text that centers on the connection between the reactive attitudes and personal relationships. And, it, too, already contains appeals on both side to empirical evidence of a great variety.

[xxii] See, for example Strawson (1962/2004).

[xxiii] A closely related argument that in one way is less and in another more, difficult to defend than the one in the text is one that begins with a distinct premise about the nature of friendship: (1) friendship necessarily generates special obligations or duties, (2) special obligations, being obligations, require the freedom to meet them. Therefore, (3) friendship requires a certain kind of free agency. This special-obligations-based argument contrasts with the one in the text: (1') friendship (or a special kind of friendship) entails the mutual recognition and appropriate holding to standards, (2') appropriately holding to standards entails free agency, (3') friendship (or a special kind of friendship) requires free agency. Interestingly, while (1) may be more widely discussed as expressing a truth about friendship than (1'), I believe that (2') is more widely assumed than (2). A fuller comparison and evaluation of both arguments awaits another occasion.

[xxiv] See Pereboom (2009), section 6, for a direct response to this claim of Kane's.

[xxv] There is, however, an appealing view that the special obligations of particular relationships only exist where there is *voluntary* assumption of them. (See, for example, Jeske (2008).) The relationship between voluntarism about special obligations and free agency is one I set aside here, however, in order to focus on what engaging in special relationships presupposes. It may be that there are multiple grounds for thinking that certain kinds of special relationships require free agency (e.g., in the free assumption of obligations necessary to relationships and in the appropriate holding to standards within relationships), but it is also possible to consider these grounds separately.