

How can psychology contribute to the free will debate?*

Shaun Nichols
Department of Philosophy
University of Arizona

1. Background: Three projects

Are people free and morally responsible? Or are their actions determined, i.e. inevitable outcomes of the past conditions and the laws of nature? These seem fairly straightforward questions, but it is important to distinguish 3 different dimensions of the free will debate: a descriptive project, a substantive project, and a prescriptive project. In this chapter, I'll consider how psychology can contribute to each project in turn. First, I should say a bit more about the projects.

The goal of the descriptive project is to determine the character of folk intuitions surrounding agency and responsibility. By uncovering the folk intuitions, one hopes to be able to sketch out the folk theory that underlies these intuitions.¹ Of particular interest for the free will debate is whether the folk notions of choice and moral responsibility are consistent with determinism. *Incompatibilists* maintain that our conceptions of free will and moral responsibility are at odds with determinism. *Compatibilists* deny this and insist that our notions of free will and moral responsibility are consistent with determinism.²

The goal of the substantive project is then to determine whether the folk views are correct. Given the folk concepts and the way the world is, does free will exist? Are people morally responsible? *Libertarians* maintain that we do have indeterminist free will (e.g. Kane 1996, O'Connor 1995, Campbell 1957). *Eliminativists* about free will maintain that free will doesn't exist. The best known version of this view is 'hard determinism', according to which we lack free will because determinism is true. However, many free will eliminativists maintain that even if determinism is false, we still lack the kind of indeterminist choice that is required by the folk notion (e.g. Pereboom 2001, Sommers 2005, Strawson 1986). On this view, our notion of free choice is incompatible with the facts, regardless of whether determinism is true or false.³

* Acknowledgements: This paper has benefited from helpful comments from John Doris, Josh Greene, Tamler Sommers, and Manuel Vargas.

¹ "Theory" is intended in a weak sense here. Roughly, any internally represented body of information will count as a theory in the intended sense.

² Some maintain that compatibilism is true with respect to responsibility but not with respect to free will (see Fischer 1999).

³ Free will eliminativists embrace the strong claim that there is no such thing as free will. An alternative would be to maintain that while we are mistaken in various of our beliefs about free choice, this doesn't mean that our concept of free will is empty (for discussion, see Vargas 2005). On such a view, one might say that just as we mistakenly thought that whales were fish, we also mistakenly thought that free will required indeterminism. In

The prescriptive project is different from both the descriptive and substantive projects. For here the question is how we should react, given what we know about our concepts and the world. Should we revise or preserve our practices that presuppose moral responsibility, like practices of blame, praise, and retributive punishment?

It might be helpful to summarize with a tree diagram the philosophical positions that emerge from these questions (see Figure 1).

Insert figure 1 about here

Let's start at the top with the central descriptive question, "Is the folk concept of free choice compatible with determinism?" If the answer is 'yes', then *compatibilism* is the right view, and at that point, at least as far as the free will/determinism issue goes, we need not bother with the substantive and prescriptive matters. If our concept of free will is happily consistent with determinism, then determinism poses neither a substantive nor a moral threat to our current views and practices (Hume 1955 [1743]). However, if the answer at this juncture is 'no', then *incompatibilism* is the right view (Kant 1956 [1788], Reid 1969 [1788], D'Holbach 1970 [1770]). If so, we face a pressing substantive question: "Does the folk concept of free choice reflect the nature of choice?" In particular, it becomes a major concern whether determinism is true. The view that choices are *not* determined is *libertarianism*. Libertarians typically hold that on the folk view, choices are not determined, and also that the folk view of choice matches the way the world is. If libertarianism is right then the prescriptive question is not pressing. For our normal practices of regarding people as free and responsible are perfectly appropriate – people *are* free and responsible. According to *free will eliminativists*, however, the answer to the substantive question is that we lack the kind of free will we think we have. On their view, the facts about the world are at odds with the way we think of ourselves. If this worrying view is correct, then the prescriptive question takes on great significance. If our folk view of choice is wrong, what is the appropriate response? Should we stop treating each other as free and morally responsible agents? *Revolutionism* is the view that we should overhaul our practices that presuppose free will and moral responsibility. *Conservatism* is the view that we should leave practices more or less untouched. Now let's see what psychology can do for each of these projects.

2. The descriptive project

Many philosophical issues are resolutely technical and detached from lay commitments. If you tell undergraduates that some philosophers, the Logicians, think that mathematics can be derived from logic, the typical response will be "who cares?" Most people simply don't have intuitions about the logical foundations of mathematics. This is not what happens when students are told that some philosophers maintain that every decision a person makes is an inevitable consequence of what happened prior to the decision. That is a disturbing suggestion even to students whose only interest in Logicism is whether it will be on the final exam. The problem of determinism and free will strikes a deep worry in us. The fact that the problem of free will resonates with people is a *psychological* fact. The descriptive project strives to capture the nature of our response to the problem of free

both cases we were wrong – whales aren't fish and free will isn't indeterministic – nonetheless there are whales and there is free will.

will – what is it that we think about free will and responsibility? Is our notion of free choice incompatible with determinism? Where does our notion of free will come from? Psychology is obviously in an excellent position to help answer these questions.

2.1. The folk notion of choice

Do the folk have libertarian, i.e., indeterminist, views about choice? This issue has only recently been explored empirically. Over the last few years, there has been a modicum of evidence that we do have indeterminist views about choice. Recent experiments investigated whether children think that agents could have done otherwise than they did (Nichols 2004a). Children were placed in one of two conditions: those in the *agent* condition witnessed an agent exhibit motor behavior; those in the *object* condition witnessed an object move. For instance, children in the agent condition were shown a closed box with a sliding lid; the experimenter slid the lid open and touched the bottom. Children in the object condition were shown the closed box with a ball resting on the lid; the experimenter slid the lid open and the ball fell to the bottom. The child was asked whether the agent/object had to behave as it did after the lid was open, or whether it could have done something else instead. The results were very clear. Every single child said that the person could have done something else and nearly every child rejected this option for the thing. In a second study, adults and children were asked about physical events, e.g., a pot of water coming to a boil, and moral-choice events, e.g., a girl stealing a candy bar. Participants were asked whether if everything in the world was the same up until the event occurred, the event had to occur. In this setting, both adults and children were more likely to say that the physical events had to occur than that the moral choice events had to occur. This provides preliminary evidence that the folk have a concept of free choice on which agents *could have done otherwise*.

Further support for the claim that people regard choice as indeterminist comes from recent experiments that Joshua Knobe and I have conducted (Nichols & Knobe forthcoming). We presented subjects with a questionnaire that depicted both a determinist universe (A) and an indeterminist universe (B), described as follows:

The key difference... is that in Universe A every decision is completely caused by what happened before the decision – given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does not have to happen* the way that it does.

After this description, subjects were asked, “Which of these universes do you think is most like ours?” The vast majority of subjects answered that the *indeterminist* universe (Universe B) is most like ours. Note that the only feature of the universe that is indeterminist is choice. So, the responses indicate that people are committed precisely to the idea that choice is indeterminist.⁴

⁴ Even if this is right, it would be premature to conclude that people are consistently indeterminist about choice. For different kinds of questions seem to provoke determinist responses (Nichols 2006). Given the central goal of this chapter, I will set aside this important complication.

The above work merely represents one view (for a contrasting account, see Nahmias et al. 2006; Nahmias forthcoming). Obviously further work would need to be done to confirm that people have indeterminist views about decisions. Even apart from this controversy, there are important additional questions that are ripe for psychological exploration. Perhaps the most interesting question is whether the results on folk indeterminism would extend to other cultures. In light of the findings on cross-cultural differences in intuitions and attitudes (see, e.g. Nisbett et al. 2001; Machery et al. 2004), it seems quite possible that the preliminary findings of indeterminist views about free choice will be restricted to Western culture. On the other hand, it might be that the belief in indeterminist free will is, as some philosophers have maintained, a deep feature of creatures like us (see, e.g. Strawson 1986; Nagel 1986).

2.2. Where does the belief in libertarian free will come from?

Assuming we have a notion of libertarian free will, a new question emerges for the descriptive project: how did this notion come about? Several accounts of the origin of our belief in libertarian free will has been offered. None of the accounts, however, has achieved adequate support. I'll review some of these accounts and their evidentiary shortcomings, then I'll suggest a different, but also unsupported, account of the belief.

The traditional explanation for how we come to believe in indeterminist agency is that it comes from introspection (e.g. Reid 1969 [1788], Holbach 1970 [1770]). For introspection fails to reveal any deterministic underpinnings of my decision making. Of course, both libertarians and determinists can agree to this. The determinist maintains that we fail to introspect the deterministic processes that actually produce our behavior. The libertarian will insist that there is no such deterministic process, so of course we don't introspect it.

Although the fact that we don't introspect deterministic causes of our choices is almost certainly part of the story, it can hardly be a complete explanation of how we acquire the belief in indeterminist choice. The fact that we don't perceive deterministic decision-making processes doesn't yet explain why we would believe that our decisions are *indeterministically* generated. For we often think processes are deterministic even when we don't perceive deterministic causal transactions.⁵ Even some of our behaviors have this quality. When my eye twitches, I have no idea what causes it, but this doesn't remotely lead me to think that eye twitches are indeterministically generated. Thus the fact that we don't perceive a deterministic process of decision making must be supplemented to explain the intuition that our decisions are not determined. The natural supplement would be to maintain that we have a standing belief that we do have introspective access to all the causal processes underlying our own decision making. If we have a standing belief in such introspective transparency and we fail to introspect deterministic processes, we might infer that there are none.

This combination of presumed introspective-access coupled with introspective failure might provide an explanation for how we come to believe in our own indeterminist agency. But there are several shortcomings. First, we would need evidence that people do in fact have a standing belief that we have introspective access to all the

⁵ Indeed, on some views, we *never* perceive the causal powers that we presume to underwrite deterministic processes.

causal processes underlying our own decision making. Second, the account also has to assume that people carry out the required inference to arrive at the view that agency is indeterminist. Third, this account would only apply to oneself. It would need to be supplemented to explain why we think that *other* people's decisions are indeterministic. None of these shortcomings is decisive, of course. But clearly psychology has an important role to play in investigating the promise of this account.

Another tempting account of the genesis of the belief in free will is that it is inferred from seeing creatures exhibit spontaneous motion. That might spawn the idea of freely generated action. Joshua Greene & Jonathan Cohen (2004) make some suggestive remarks along these lines. First, they note, in keeping with much work in developmental psychology (e.g. Baron-Cohen 1995; Leslie 1995), that the mind plausibly has different systems for dealing with matter and for dealing with minds. The system for dealing with minds is triggered by a wide variety of stimuli including the Heider/Simmel animation of spontaneously moving geometric shapes (e.g., Heider & Simmel 1944). Greene and Cohen link their proposal to such findings and maintain that we regard others as free because we regard them as having minds. They write, "we suggest that a crucial feature, if not the defining feature, of the mind (intuitively understood) is that it's an *uncaused causer* (Scholl and Tremoulet, 2000). Minds animate material bodies, allowing them to move without any apparent physical cause and in pursuit of goals" (35, emphasis added). On their view, ordinary objects like rocks seem to obey ordinary physical laws – "these things don't get up and move around on their own." There are other things however that "seem to operate by some kind of magic... [moving] about at will, in apparent defiance of the physical laws that govern ordinary matter" (32). This suggests the following sort of account: spontaneous motion triggers the attribution of a mind as a free agent – an uncaused causer. Although Greene and Cohen don't endorse quite this picture, it is an attractively simple hypothesis about why we believe in free will.

Attractive, but mistaken. First, it is quite possible that the system for dealing with predicting and explaining minds is *deterministic*. Indeed, in the vast literature on "mindreading", the notion of indeterminist choice is not invoked in any of the prevailing models of how children predict and explain behavior (see, e.g., Gopnik & Meltzoff 1997, Leslie 1995, Nichols & Stich 2003). That is, none of these models maintains that people routinely invoke the notion of indeterminist choice when they predict and explain behavior. This makes sense, since the predictive aim of the mindreading system would not be improved by the inclusion of an assumption of metaphysical indeterminism. So, we cannot assume that we get attributions of libertarian free will as a freebie once we have attributions of minds. Second, even babies apparently attribute mental states to various phenomena, including certain computer-generated geometric objects of the Heider/Simmel sort adverted to by Greene & Cohen (e.g. Csibra 2003; Kuhlmeier et al. 2003; Premack 1990). Recall that to attribute libertarian free will to an agent is to maintain that the agent could have done otherwise even if everything else had been the same; that seems a more sophisticated thought than many of us are willing to impose on the baby. Finally, recent evidence from Susan Johnson and colleagues suggests that spontaneous motion is not sufficient to activate attributions of mental states.⁶ In one experiment, 12 month old infants were shown a fuzzy brown object under a variety of

⁶ See also Woodward (1998).

different conditions (Johnson, Slaughter & Carey 1998). In one condition, the fuzzy brown object interacted contingently with the infant by beeping and flashing when the infant babbled or moved; in another condition, the fuzzy brown object exhibited an equivalent amount of apparently self-generated flashing and beeping, but in this condition the activity was not contingent on the infant's behavior. In both conditions, children's looking behavior was measured when the fuzzy brown object 'gazed' at one of two objects by making a smooth, 45 degree turn towards the object and remaining in this orientation for several seconds. What Johnson and colleagues found was that infants would follow the "gaze" of the fuzzy brown object when its spontaneous activity was contingent, but not when the spontaneous activity was noncontingent. Johnson and colleagues propose that what happened in the experiment is that the infants followed the gaze when the fuzzy brown object was coded as an *agent*. For gaze-following is often taken to reflect the "implicit attribution of a mind to the gazer" (Johnson et al. 1998, 233).⁷ This suggests that mere self-generation of behavior isn't sufficient for attribution of a mind. For in the noncontingent condition, infants don't seem to attribute mental states despite the presence of spontaneous motion. Of course, it's possible that babies to attribute minds and free will to these spontaneous movers, and then quickly revise their beliefs. But without evidence to that effect, the spontaneous motion account does not yet explain the attribution of free will.

In Nichols (2004a) I suggested, *very tentatively*, an alternative account on which the acquisition of the belief in indeterminist choice derives from a prior belief in obligation. According to a famous Kantian argument, we can prove that we have indeterminist choice from the maxim "*ought* implies *can*" and the fact that we ought to follow the moral law. The idea is that we can't be obligated to do the impossible, and if determinism is true, it is impossible for us ever to do other than we are determined to do. Thus, if we say that a person *ought* to have behaved differently, this implies that the person *could have done otherwise* (in an indeterminist sense). The suggestion in Nichols (2004a) was that, despite the dubiousness of the Kantian argument as a *proof* of indeterminist choice, it might provide an account of how we come to *believe in* indeterminist choice. There is plenty of evidence that even young children think that people *ought* to behave in certain ways (e.g. Nichols 2004b, Nucci 2001). Indeed, the child applies notions of obligation in a variety of contexts including contexts of moral transgressions (you shouldn't kick people), conventional transgressions (you shouldn't eat steak with your hands), and even simple cases of advice (you should put on sunscreen). If children apply some notion of obligation that carries the Kantian implication *could have done otherwise* (in an indeterminist sense), then the child has the essential ingredients for coming to believe that decisions are not determined.

Unfortunately, while there is abundant research showing that children apply obligation concepts, there is no evidence yet confirming the idea that children embrace the Kantian maxim. It is likely that children embrace *some* kind of ought-implies-can view. If you ask whether it was wrong for the paraplegic not to swim to save a drowning victim, children will presumably say that it's not wrong because he *couldn't* swim. But it

⁷ Johnson's research program has confirmed and extended these results in several ways (see Johnson 2003).

will be harder to show that children think that obligations carry the implication of *indeterminist*-can, and that this leads to their belief in libertarian free will.

My own newly favored view about acquisition is that the notion of indeterminism plays a larger role in childhood cognition than previously recognized (Nichols 2006). Even apart from choice events, the child often hears and uses a notion of possibility that is naturally interpreted as in conflict with determinism. Here are some examples drawn from the CHILDES database in which children advert to possibilities⁸:

Father: You could fall and get hurt Ross.

Ross (4;2): No. Not if I hold on to here and here I won't.

Father: You could... It's dangerous (MacWhinney & Snow 1990).

Ross (2;7): Marky [a younger sibling] might fall (MacWhinney & Snow 1990).

Adam (4;2): Paul [a younger sibling], you might cut yourself on this (Brown 1973).

It's natural to interpret these uses of *possibility* as in conflict with determinism. That is, it's natural to read Ross as saying that it's a genuinely open possibility that Marky will fall (or not fall), and Adam as saying that it's a genuine possibility that Paul will cut himself (or not). There are various determinist-friendly ways to read the language of possibility, and perhaps one of these ways is the best interpretation of children's modal language. But it's worth noting that some of most familiar ways to tame modality in philosophy don't look at all promising. One way to preserve a notion of *possibility* while being neutral about determinism is to treat *possibility* as a deflationary kind of epistemic possibility, on which to say that *p* is possible is to say "For all I know, *p* will happen". But this seems an implausible interpretation. When Adam says that Paul might cut himself, it seems unprincipled to maintain that Adam really just means "As far as I know, Paul will cut himself." Similarly, when Ross' father says "You could fall" and then repeats, "You could... it's dangerous", it's doubtful that Ross would interpret his dad as merely reporting on epistemic possibility. The simple epistemic possibility interpretation is even less plausible when we move to statements about past possibilities. Parents say things like "you could have broken the lamp!" And kids come to use language this way as well. In CHILDES we find Ross (at age 5) saying that he climbed on a shelf and "It could have fell on us." (MacWhinney & Snow 1990). Obviously he doesn't mean "For all I know it fell on us." Thus, the simple epistemic gloss fails to provide a general account of children's judgments of possibility.

If the child does have an indeterminist notion of *possibility* outside of choice contexts, we get a new avenue for exploring the acquisition of the concept of free will. For now it seems that the child's indeterminism about choice is part of a more general indeterminism about possibilities. Elsewhere I've suggested that a primary function of the modal concept POSSIBLE is to represent risk and opportunity (Nichols forthcoming). This was based on an informal review of modal talk in the CHILDES database. It is a

⁸ The CHILDES database is a collection of transcripts of spontaneous language use by several children in family settings.

salient fact about everyday conversation that children and parents tend to use modal language largely to convey information about risks and opportunities. We've already seen some examples that represent risks. Here are some examples of children deploying modals to represent opportunities:

Alison (2;4): We could march around or we could run around (Bloom 1973).

Ross (3;3): Hi Titus [a cat]. I got her tail.

Father: You did.

Ross: She's under there. I could get her.

Father: Okay but don't be too mean to her okay? (MacWhinney & Snow 1990).

Father: I can't make that one work.

Abe (3;7): You could glue it (Kuczaj & Maratsos 1975).

Adam (4;1): We could put the animals in here (Brown 1973).

Adam (4;10): Ursula you could stay and eat with us (Brown 1973).

And here are remarks about opportunities made to children by adults.

Mother: Adam you could draw some pictures now (Brown 1973).

Mother: you could shoot at anything that's set up for a target (Brown 1973).

Father: you could go visiting to other people's houses that'd be fun (MacWhinney & Snow 1990).

Mother: you could play with your birdies honey (Sachs 1983).

As parents, when we point out risks to our children, what matters to us is making sure that our children are safe, that they avoid unnecessary dangers. We typically don't care about trying to communicate a carefully qualified notion of possibility that is neutral about determinism. Similarly, when we point out opportunities to our children, we aren't concerned to get them to hedge the modal notions in a compatibilist way. We want to stress the options before them. In short, when we alert our children to risks and opportunities, deterministic explanation is pretty much the last thing on our minds. Or theirs. In this light, it should not be surprising if our notion of possibility fails to be nuanced in a compatibilist fashion. The concept of *possibility* can serve the key function of representing risks and opportunities quite well without any compatibilist subtlety.⁹

Although the above account seems most plausible to me at the moment, the lay understanding of modality is dramatically underexplored. In particular, there is precious

⁹ Of course if we had access to deterministic causal mechanisms underlying our decisions, we would not think them indeterminist. But the same goes for risks. When we have access to deterministic causal mechanisms, we don't think the outcome was indeterminist.

little experimental evidence on the extent to which children apply an indeterminist notion of *possibility*. Once again, psychology is uniquely well positioned to confirm or disconfirm the acquisition story.

3. The substantive project

Substantive questions about free will and determinism – are our choices determined? – are most pressing for incompatibilists. For if free will is incompatible with determinism, then it is imperative that we know whether our actions are determined. The arguments in favor of libertarian free will typically appeal to introspection (e.g. O'Connor 2000, Campbell 1957). Since few psychologists would countenance such arguments, we will consider here only arguments against libertarianism.

3.1. *A priori* arguments

A number of philosophers argue on *a priori* grounds that libertarian free will is incoherent or impossible (e.g. Double 1991; Strawson 1986, 1994). Perhaps the most enduring *a priori* critique is Hobbes' libertarian dilemma. On the one hand, libertarians say that an agent's decision isn't free if the decision is determined. But on the other hand, if the decision is *not* determined, then it isn't determined by the agent either! That, critics maintain, leaves the libertarian fresh out of intelligible options (for discussion see Kane 1996, 11). As the libertarian dilemma illustrates, libertarian free will can seem decidedly mysterious and counterintuitive, and this counts as a strike against it. But, as evidenced by quantum mechanics, some mysterious and counterintuitive views are apparently true. So it would be overly hasty to conclude from such *a priori* arguments that libertarian free will is metaphysically impossible.

3.2. *A posteriori* arguments

In any case, whatever one wants to say about the *a priori* arguments, this is not a natural place for psychology to make a contribution. A more natural role for psychology is in making *a posteriori* objections to libertarian free will. Psychology might, for instance, show that (psychological) determinism is true, thereby dealing a direct blow to libertarianism. The most explicit development of this kind of attack comes from John Bargh and Melissa Ferguson (2000). Their argument runs into serious troubles, I think, and the troubles illustrate why it will be very difficult for 21st century psychology to prove that libertarian free will doesn't exist. So I want to discuss Bargh and Ferguson's argument at some length.

Bargh and Ferguson adopt a standard notion of determinism: "for every psychological effect (e.g., behavior, emotion, judgment, memory, perception), there exists a set of causes, or antecedent conditions, that uniquely lead to that effect" (2000, p. 925), and they claim that psychology has provided evidence for determinism. For instance, they write, "behaviorists and cognitive (and social-cognitive) scientists have accumulated evidence of determinism by their many demonstrations of mental and behavioral processes that can proceed without the intervention of conscious deliberation and choice" (925). Bargh and Ferguson rightly note that the existence of consciously controlled decisions does nothing to refute determinism. But they then go on to argue for the much stronger thesis that work on control processes provides clear evidence for

determinism: “the automatic goal operation experiments provide... rather obvious evidence that even controlled mental processes are themselves controlled and determined” (939). As an example of such evidence, Bargh and Ferguson advert to an important experiment by Chartrand & Bargh (1996). In this experiment, participants first completed a “scrambled sentence” task in which they are told to form grammatical sentences out of short lists of words. In one condition (impression goal condition), the task was loaded with terms associated with impression formation (e.g. ‘evaluate’, ‘judge’, ‘assess’); in the other condition (memory goal condition), the task was loaded with terms associated with memory (e.g. ‘retain’, ‘remember’). All participants then read several sentences describing the behavior of a given person, and were told that they would be asked about this later. Strikingly, participants in the impression goal condition had more accurate recall than the other participants. These participants also showed better organization of the information in memory. This effect conforms to earlier results in which subjects are explicitly instructed to inform an impression or to memorize. What Chartrand & Bargh show is that even when the goal is induced implicitly, the effect still emerges. Indeed, in their experiment, participants in the two conditions were unaware that the goal had been induced – the two groups showed no difference when asked what they were trying to do when they read the descriptions of the person’s behavior (Chartrand & Bargh 1996).

There are a number of problems with using this evidence to try to undermine libertarian free will. First, libertarians maintain that determinism is false about *decisions*. Libertarians have typically been willing to allow that determinism is true about other psychological processes. Libertarians take considered *moral* decisions to be a paradigm example of the kind of mental activity that is a good candidate for not being determined. Strategy formation of the sort demonstrated by Chartrand & Bargh does not fit this paradigm very closely.

Although this might be a problem with using this particular bit of evidence to support determinism, I suspect that the ever ingenious John Bargh could generate evidence that even moral decisions are affected by nonconscious goals. But there is a deeper with problem with the argument. The results are, of course, statistical. What they show is that nonconscious goals *influence* psychological outcomes. That’s a far cry from showing that the psychological outcomes are *determined*. To make this point as plain as possible, consider another delightful result from Bargh’s lab (Bargh, Chen & Burrows 1996). Again, subjects were given a scrambled sentence task. This time participants were assigned to different conditions in which the word sets were loaded with terms related to rudeness, to politeness, or to neither. All participants were told that after completing the scrambled sentence task, they were to go tell the experimenter that they were ready for the next task. However, the experimenter was engaged in conversation, and the participant would have to interrupt in order to tell the experimenter. Bargh and colleagues found that among those primed for rudeness 63% interrupted, among those primed for politeness only 17% interrupted, and for the control group, 37% interrupted. Again, subjects didn’t subsequently explain their behavior by invoking exposure to terms of politeness or rudeness. This is a remarkable demonstration that our decisions are influenced in ways that fall well outside of our awareness. But obviously there’s still lots of individual variation here. The politeness prime didn’t eliminate interruptions, and the

rudeness prime didn't send interruptions to ceiling. A libertarian can perfectly well maintain that indeterminist free choice accounts for some of the variance in these studies.

A more general point can be made now. Virtually all libertarians are happy to allow that there are many factors that influence our choices (Campbell 1957, Clarke 1993, Kane 1996, O'Connor 1995). Our natural inclinations, genetic predispositions, and involuntary appraisals all influence outcomes. The terrific work in social psychology shows that there are hitherto unrecognized influences. But they are still just more factors that the libertarian can happily concede. The difficulty in excluding indeterminist choice is that one would need evidence of something stronger than influence. We would need to show that the decision was *entirely* produced by a known set of influences. As far as I know we don't have a single worked out case of the deterministic processes that generated a single decision of a single individual. Thus, we are not in a good position to claim that determinism has been demonstrated.

The challenge for the *a posteriorist* is actually even greater. For even if we found that many decisions are determined, this still wouldn't refute libertarianism. A libertarian can perfectly well maintain that libertarian free activity is a relatively rare phenomenon. Indeed, one prominent scientific libertarian, Robert Kane, maintains that even my "free" acts can be determined, so long as certain determining factors have their ultimate source in a "self forming" event. Even though they might be rare, these self-forming events are the key to our ultimate freedom and responsibility according to Kane (1996, 75-8). Put simply, the problem for *a posteriori* arguments for determinism is that determinism is a universal claim – every event is determined. Libertarians, by contrast, don't think that every event is indeterministic. Indeed, some libertarians don't even think that every choice event is indeterministic. Thus, to establish determinism on the basis of psychological evidence would require evidence that suffices for a universal generalization. That's a towering order. At best, it will be a long time before the psychological sciences exclude the rare occurrence of indeterminist free will.

While psychology is not currently equipped to prove that determinism is true, it is plausible that determinism is an important guiding principle for psychological research. This is also part of Bargh and Ferguson's brief. They write "psychologists studying higher mental processes should continue the scientific study of conscious processes but at the same time give appropriate attention to the deterministic philosophy that must underlie such analysis" (2000, 940). Here, they are on much better ground. I do think that psychological determinism has been and will continue to be a vital assumption guiding research. And I'm inclined to think it's true. But my allegiance was not produced by checking the evidence. Rather, it came from an abiding conviction that people's decisions *have* to have an explanation.

3.3. Genetic arguments

Even if the *a priori* and *a posteriori* arguments against libertarian free will are limited, there is another way that psychology can contribute to the substantive project. Rather than arguing that libertarian free will doesn't exist, we might argue that our belief in libertarian free will is unjustified. Freud used genetic considerations to argue that religious belief is unjustified. On his view, when we see the source of our religious beliefs, we will come to appreciate that they are unwarranted (Freud 1961 [1927]).

Similarly we might find that the source of our belief in libertarian free will reveals that the belief is unwarranted.

In section 2.2, I reviewed several different accounts of how we come to believe in libertarian free will, and at least in some cases, if the accounts are right, then it seems that we are not justified in our belief in libertarian free will. To see this, let's consider the last account that I reviewed in section 2.2, coupled with the idea that we lack introspective access to any deterministic processes. I suggested that our belief in libertarian free will is part of a more general indeterministic outlook that applies to risks and opportunities. Parents present risks and opportunities as genuine, indeterministic, possibilities, and children (and adults) represent them as such. Since the purpose of these representations is to avoid danger and capitalize on opportunity, there is no direct cost in failing to hedge them in compatibilist ways. The representations function equally well regardless of whether they accurately represent indeterministic events or whether they inaccurately represent events that are deterministic but unpredictable by us.

So are these indeterministic beliefs warranted? In the case of our indeterministic beliefs about risks (e.g., "the branch might fall on you!"), the right thing to say is that our beliefs are unwarranted. We lack the discrimination to see whether natural events concerning middle-sized objects are deterministic. Indeed, our everyday observations are entirely consistent with a deterministic physics, even if more careful observation exposes indeterminism. Because of our limited powers of discrimination in everyday life, we would believe events to be indeterministic even if they aren't. As a result, when the child (or adult) says, "The branch might fall on you," he isn't warranted in thinking that it's a *metaphysically* indeterministic possibility. Rather, he is only warranted in a belief about *epistemic* possibility – *for all I know the branch will fall*.

Similarly, we lack the discrimination to see whether the choices people make have a libertarian source. Obviously we lack the necessary discrimination when it comes to other people – we have no direct access to the psychological processes that eventuate in their decisions. But research in social psychology suggests that even introspection fails to provide the kind of access that would be needed to detect whether our own choices are generated by libertarian free will. For introspection fails to detect the causal influence of many causal factors, including both conscious and unconscious mental states (e.g. Bargh & Chartrand 1999, Nisbett & Wilson 1977, Wegner 2002, Wilson 2002). As a result, even if our decisions are deterministically generated, introspection would still fail to reveal this to us. So, our belief in libertarian free will depends crucially on mechanisms that are too insensitive to tell us whether our choices are in fact generated by indeterministic free will. *Prima facie*, then, our belief in libertarian free will lacks any good rational foundation. Note that this is not to say that the evidence on introspection shows that we lack libertarian free will; rather, the evidence helps to show that our belief in libertarian free will is not well grounded. Thus, while we might not have compelling empirical evidence against libertarian free will, we lack good reasons to trust our belief in libertarian free will.

It's useful to contrast this epistemic situation with our position with respect to quantum mechanics. As noted in section 3.1, libertarian free will is weird – it's hard to devise a clearly intelligible and intuitive account of libertarian free will. But the same can be said for quantum mechanics. Quantum mechanics is so deeply weird that there's some question about whether we have the cognitive resources to understand it. So what's

the difference between quantum mechanics and libertarian free will? The difference is that we came to believe in quantum mechanics on the basis of mathematical proofs and experimental evidence. Our belief in libertarian free will lacks any such impressive credentials. On the contrary, our enduring belief in indeterminist free will, like our belief in indeterminism for middle-sized objects, can easily be explained by our lack of discrimination.

This sort of genetic argument depends on certain philosophical assumptions about when we lack justification. But more importantly for present purposes, the argument depends on a speculative story about how we come to believe in libertarian free will. We need a well confirmed psychological account to develop any such argument adequately. But I suspect that if we want to know in our lifetimes whether we should believe in libertarian free will, our best hope is a psychologically-informed genetic argument.

4. The prescriptive project

When we turn to the prescriptive project, the question is, *what should we do?* On this question, one might expect that psychology has absolutely nothing to contribute. After all, this is a question about *ethics*. However, I think psychology might make major contributions even here. For if knowing the facts will help us make better decisions about what we should do, the facts do make a contribution to prescriptive concerns. As noted in section 1, the prescriptive project is especially pressing for free will eliminativists. For on that view, there is a fundamental mistake in lay notions of free will and responsibility. Two key questions that then emerge for the eliminativist are: (i) what *would* happen if people stopped believing in libertarian free will and (ii) would a revolution be on balance beneficial? Psychology is poised to help with each question.

Some philosophers have worried that if people give up on free will, this might have dire effects on everyday life, and as a result, some have suggested that we ought to keep the truth hidden from the *hoi polloi*. This is actually a venerable strand of worry in ethics. A number of philosophers who embrace a utilitarian ethics maintain that there would be bad consequences if the man on the street actually knew that utilitarianism captured the truth about ethics. Hence, they counsel concealment. In his influential utilitarian treatise, Henry Sidgwick writes:

...a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands (1907: 490).

More colorfully, William Lycan writes “I believe ... firmly in some form of act-utilitarianism in ethics, but the sacred principle of utility itself forbids me even telling you this” (Lycan 1987, 136n1). The worry is that people will behave badly if they come to believe utilitarianism. Hence, the utilitarian maintains that we should keep the truth secret.

Similarly, if we thought that anarchy and despair would ensue if people knew that there is no libertarian free will, this might count as a reason to resist informing the public of the truth. Such in fact is the view of Saul Smilansky. He writes, “Humanity is

fortunately deceived on the free will issue, and this seems to be a condition of civilized morality and personal value” (2002; 500), and “there would be considerable room for worry if people became aware of the absence of libertarian free will” (2000, 505n7).

If people become convinced that they lack libertarian free will, would terrible consequences ensue? This is clearly an empirical question. And the rather limited empirical evidence suggests that we needn't fear a catastrophe. A recent experiment indicates that people's views about responsibility and determinism shift depending on whether determinism is presented as a remote possibility or a very live possibility. In the experiment, subjects read non-technical descriptions of determinism. In one condition, subjects were told to imagine another universe that was deterministic, and in the other condition, subjects were told to imagine that *our* universe was deterministic. In the “other universe” condition, subjects tended to say that it's impossible for agents to be fully morally responsible in that universe, however in the “our universe” condition, subjects tended to say that it would still be possible for agents to be fully morally responsible in our universe (Roskies & Nichols forthcoming). Thus, it seems that when people consider determinism as a genuine possibility for our world, they tend not to descend into anarchic moral views. Research using a quite different methodology also suggests that if people became determinists, it would have little effect on their judgments of moral responsibility. Viney and colleagues found that college students who were identified as determinists were no less punitive than indeterminists (Viney et al. 1982) and no less likely to offer retributivist justifications for punishments (Viney et al. 1988). Although these results from Viney and colleagues are suggestive, the measure used for identifying determinists is flawed, and so here is another obvious place for further psychological research.

The final and deepest prescriptive question asks what we should do, all things considered, if we lack libertarian free will. Obviously this question is so broad that I can scarcely do it any justice, but let's consider one social concern that looms large in the free will debate – retributive punishment. Retributivist approaches to punishment, which are central to our penal system, maintain that some people *deserve* to be punished because they are *guilty*. Several free will eliminativists have promoted a revolution in the penal system, suggesting that given the absence of libertarian free will, we ought to stop the practice of retributive punishment (e.g. Greene & Cohen 2004; Pereboom 2001). But it seems rash to try to overhaul these practices before we have some idea about the consequences of such a revolution. If Marxists had been more attentive to the psychological facts about human nature, we might have been spared some disastrous experiments in social engineering.

Once again, the natural place to look for insight is psychology. Retributive punishment comes naturally to us. This is nicely illustrated in a study by Haidt and Sabini (2000) in which they showed subjects film clips that depicted injustices, and then subjects were asked to rate different possible endings. Subjects were not satisfied with endings in which the victim dealt well with the loss and forgave the transgressor. Rather, subjects were most satisfied when the perpetrator suffered in a way that paralleled the original injustice. That is, subjects preferred the ending in which the perpetrator *got what he deserved*. Recent work in experimental economics indicates that these motives for retributive punishment might play an important positive role. Fehr & Gächter (2002) found that people will pay to punish those who defect in public goods games even when

there is no reason to think that such punishment will produce benefits for themselves. This kind of punishment seems retributive in nature – as in Haidt & Sabini’s study, the subjects here presumably think that the defector deserves to be punished (as opposed, say, to thinking that it will produce better utilitarian outcomes). Fehr & Fischbacher (2004) showed that punishment of this sort is extremely effective at establishing cooperation. In one such demonstration, participants play a public goods game in which they have the option to cooperate by contributing significantly to a common fund or defect by not contributing. In the first 10 trials, there is no opportunity to punish others, and contributions drop precipitously across these 10 trials. Then subjects are informed that they will be able to punish others in subsequent trials. Immediately after they receive this information, the level of cooperation takes a great leap, and after a few trials, the level of cooperation is about 90% of the maximum possible (Fehr & Gächter 2000). Thus, the presence of retributive punishment can secure cooperation, and merely knowing that others are in a position to exact retribution makes a person more likely to cooperate.

The foregoing represents only one line of research on punishment. But if the role of retributive punishment is as important as these studies suggest, then to promote the abolition of retribution seems a dangerous cause. It is especially important to recognize that retributive punishment is so effective partly because we are so receptive to it. Merely knowing about the availability of retributive punishment seems to improve people’s behavior. Of course, we need to consider as well the potential costs of keeping retributive punishment. But it would be wise to wait for more evidence before fomenting a revolution.

5. Conclusion

In this chapter, I’ve sketched the complex intellectual geography of the free will debate. This geography includes three distinct dimensions. On the descriptive dimension we want to discern the character of lay views concerning free will and the origin of those views. On the substantive dimension, we want to know how well the lay views capture the way things really are. And on the prescriptive dimension, we want to know how we should act in light of what we find out about the existence of free will. For each dimension, I’ve argued, psychology has an important role to play. Indeed, psychology holds great promise for advancing our understanding here. Philosophers have long tried to resolve the free will debate in isolation from psychological science, and despite the enormous ingenuity and effort that has been exerted, the purely philosophical investigations have not led to any kind of wide consensus. Psychology is poised to breathe new life into these issues. And one might hope that by bringing the resources of psychology to bear on the issue of free will, we can achieve the consensus that philosophy alone has failed to reach.

References:

- Bargh, J. A., & Chartrand, T. L. 1999. The unbearable automaticity of being. *American Psychologist*, 54, 462-479.
- Bargh, J. & Ferguson, M. 2000. Beyond behaviorism: on the automaticity of higher mental processes. *Psychological Bulletin*, 126, 925-45.

- Baron-Cohen, S. 1995. *Mindblindness*. Cambridge, MA: MIT Press.
- Bloom, L. 1973. *One Word at a Time: The Use of Single Word Utterances before Syntax*. The Hague, Netherlands: Mouton.
- Brown, R. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Campbell, C. 1957. *On Selfhood and Godhood*. London: George Allen & Unwin.
- Chartrand, T. L., & Bargh, J. A. 1996. Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, 71, 464-478.
- Clarke, R. 1993. Toward a credible agent-causal account of free will. *Nous* 27: 191-203.
- Csibra, G. 2003. Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London*, B 358, 447-458.
- Double, R. 1991. *The Non-Reality of Free Will*. New York: Oxford University Press.
- Fehr, E. & Fischbacher, U. 2004. Social norms and human cooperation. *Trends in Cognitive Sciences* 8, 185-190.
- Fehr, E. and Gächter, S. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980-994
- Fehr, E. & Gächter, S. 2002. Altruistic punishment in humans. *Nature* 415, 137-140.
- Freud, S. 1961 [1927]. *The Future of an Illusion*. Translated by J. Strachey. New York: Norton & Co.
- Gopnik, A. & Meltzoff A. 1997. *Words, Thoughts and Theories*. Cambridge, MA: MIT Press.
- Greene, J. & Cohen, J. 2004. For the law, neuroscience changes nothing and everything, *Philosophical Transactions of the Royal Society of London B*, 359, 1775-17785.
- Haidt, J. and Sabini, J. forthcoming. What exactly makes revenge sweet?
- Heider, F. and Simmel, M. 1944. An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.
- Holbach, P. 1970 [1770]. *The System of Nature: Or, Laws of the Moral and Physical World*. Translated by H. D. Robinson. New York, B. Franklin.
- Hume, D. 1955 [1743]. *An Enquiry concerning Human Understanding*. L. Selby-Bigge (ed.). Oxford: Clarendon Press.
- Johnson, S. 2003. Detecting agents. *Philosophical Transactions of the Royal Society of London B*, 358, 549-59.
- Johnson, S., Slaughter, V. and Carey, S. 1998. Whose gaze will infants follow? Features that elicit gaze-following in 12-month-olds. *Developmental Science*, 1, 233-238.
- Kane, R. 1996. *The Significance of Free Will*. New York: Oxford University Press.
- Kant, I. 1956 [1788]. *The Critique of Practical Reason*. Trans. L. Beck. Indianapolis: Bobbs-Merrill.
- Kuczaj, S. and Maratsos, M. 1975. What children *can* say before they *will*. *Merrill-Palmer Quarterly*, 21, 87-111.
- Kuhlmeier, V.A., Wynn, K., & Bloom, P. (2003). Attribution of Dispositional States by 12-month-olds. *Psychological Science*, 14, 402-408.
- Leslie, A. 1995. A theory of agency. In D. Sperber, D. Premack and A. Premack (eds.) *Causal Cognition*. Oxford University Press.
- Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.

- MacWhinney, B. and Snow, C. 1990. The child language data exchange system: An update. *Journal of Child Language*, 17, 457-472.
- Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Nahmias, E. forthcoming. Intuiting freedom: A response to Shaun Nichols. *Journal of Culture and Cognition*,
- Nahmias, E., Morris, S., Nadelhoffer, T. & Turner, J. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research*.
- Nichols, S. 2004a. The folk psychology of free will: Fits and starts. *Mind & Language*, 19, 473-502.
- Nichols, S. 2004b. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford University Press.
- Nichols, S. 2006. Folk intuitions about free will and responsibility. *Journal of Cognition and Culture*.
- Nichols, S. & Knobe, J. forthcoming. Moral responsibility and determinism: Empirical investigations of folk intuitions. *Nous*.
- Nichols, S. and Stich, S. 2003. *Mindreading*. Oxford: Oxford University Press.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. 2001. Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, 108, 291-310.
- Nisbett, R. and Wilson, T. 1977. Telling more than we can know. *Psychological Review*, 84, 231-59.
- Nucci, L. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.
- O'Connor, T. 1995. Agent causation. In O'Connor, ed., *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press, 173-200.
- O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- Pereboom, D. 2001. *Living without Free Will*. Cambridge University Press.
- Premack, D. 1990 The infant's theory of self-propelled objects. *Cognition* 36, 1-16.
- Reid, T. 1969 [1788]. *Essays on the Active Powers of the Human Mind*. Cambridge, Massachusetts: MIT Press.
- Roskies, A. & Nichols, S. forthcoming. Bringing responsibility down to earth.
- Sachs, J. 1983. Talking about there and then: The emergence of displaced reference in parent-child discourse. In K. Nelson (ed.), *Children's Language*, vol. 4. Hillsdale, NJ: Erlbaum, pp. 1-18.
- Scholl, B. and Tremoulet, P. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4: 299-309.
- Sidgwick, H. 1907. *The Methods of Ethics*. London: Macmillan.
- Smilansky, S. 2000. *Free Will and Illusion*. Oxford: Oxford University Press.
- Sommers, T. 2005. *Beyond Freedom and Resentment: an Error Theory of Free Will and Moral Responsibility*. PhD Dissertation, Duke University.
- Strawson, G. 1986. *Freedom and Belief*. Oxford: Oxford University Press.
- Strawson, G. 1994. The impossibility of moral responsibility *Philosophical Studies* 75, 5-24.
- Vargas, M. 2005. The revisionist's guide to moral responsibility. *Philosophical Studies*. 125, 399-429.

- Viney, W., Waldman, D., & Barchilon, J. 1982. Attitudes toward punishment in relation to beliefs in free will and determinism. *Human Relations*, 35, 939-49.
- Viney, Wayne; Parker-Martin, Pamela; Dotten, Sandra D.H. 1988. Beliefs in free will and determinism and lack of relation to punishment rationale and magnitude. *Journal of General Psychology* 115, 15-23.
- Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Weinberg, J., Nichols, S. and Stich, S. 2001. Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429-460.
- Wellman, H. 1990. *The Child's Theory of Mind*. MIT Press.
- Wilson, T. 2002. *Strangers to ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press
- Woodward, A. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.

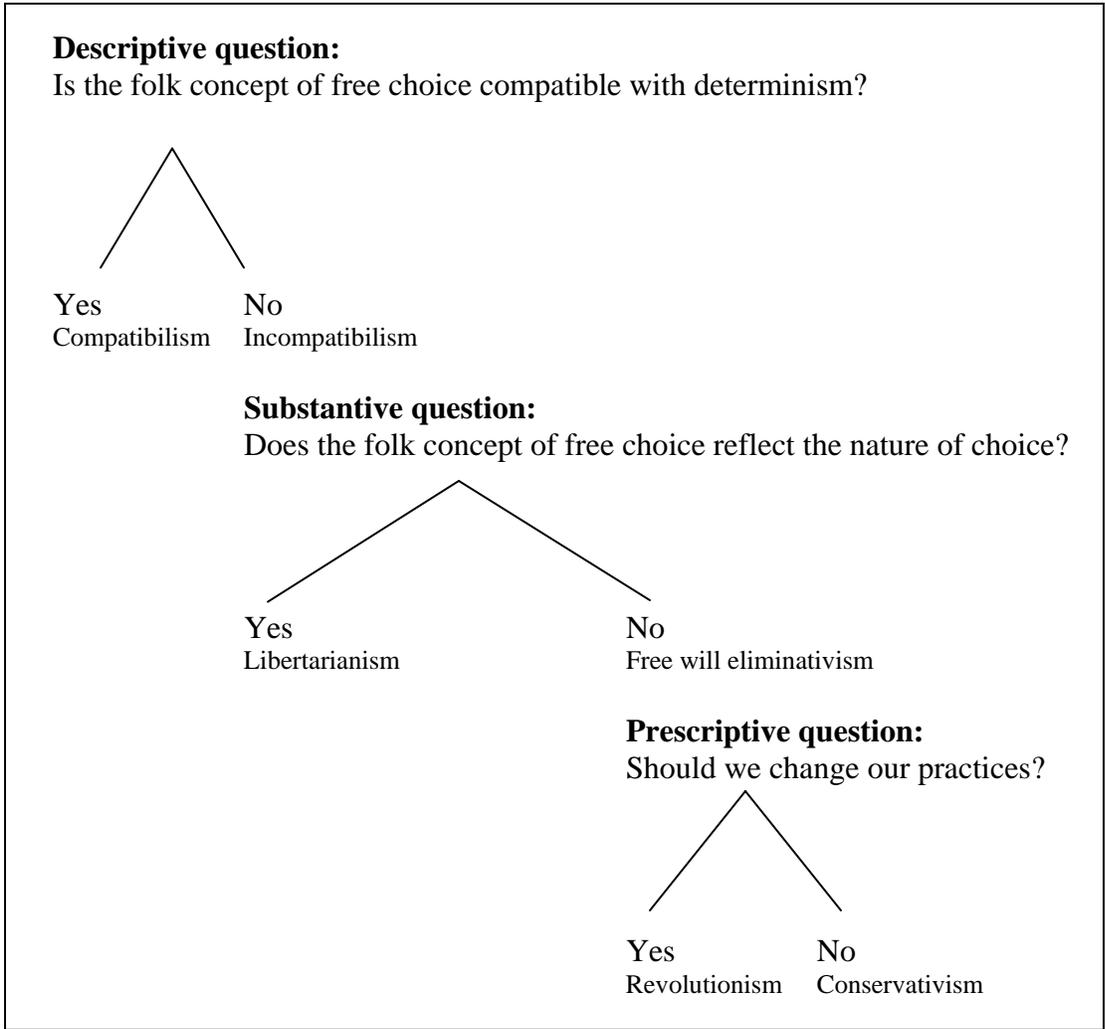


Figure 1: A taxonomy of positions organized by descriptive, substantive, and prescriptive questions