A dialogue on Free Will

JOSEPH CAMPBELL

Washington State University
josephc@wsu.edu

Much of the recent discussion concerning the problem of free will has been centered on the compatibilism/incompatibilism dichotomy. Do you think the central role attributed to this dichotomy is well deserved? And, if so, which of the two alternatives is preferable in your opinion?

Some contemporary philosophers think that the compatibilism / incompatibilism debate is dead; that we've moved on to bigger and better issues. On the contrary, I think compatibility problems are still relevant to contemporary free will debates.

Just to be clear, there are many compatibility problems: causal determinism and free will, causal determinism and moral responsibility, complete foreknowledge and free will, etc. In general, when I use the terms without qualification, "compatibilism" and "incompatibilism" convey views about determinism and free will. Thus, *incompatibilism* is the view that causal determinism entails *free will scepticism* – the view that no one has free will.

Yet many of the best contemporary arguments for incompatibilism are concerned with the compatibility of determinism and moral responsibility. In truth, these different compatibility problems – determinism/free will and determinism/ moral responsibility – have blurred in recent years since contemporary philosophers often define "free will" as the freedom relevant condition necessary for moral responsibility.

Nonetheless, some form of incompatibilism is a premise in most if not all arguments for free will scepticism. Derk Pereboom's case for free will scepticism includes a version of the *manipulation argument*, which is an argument for incompatibilism between determinism and moral responsibility (Pereboom 2001). Even in Galen Strawson's *basic argument* (Strawson (2010); Strawson (2011)), incompatibilism plays an essential role. In some versions of the argument, Strawson makes no explicit mention of either determinism or indeterminism, which leads some to think the argument is independent of the compatibilism/incompatibilism debate. Yet if you look closely it turns out that Strawson's argument takes the form of a constructive dilemma. He argues that ultimate moral responsibility is impossible by first first arguing that it is incompatible with determinism and then showing that indeterminism can't help (Campbell 2011, 1.5; 3.4; 4.3). Strawson writes: "It is, however, absurd to suppose that indeterministic factors, for which one is obviously not responsible, can contribute in any way to one's being truly morally responsible for how one is" (Strawson 2011, p.128).

My view is that if one can show that arguments for incompatibilism are faulty, or at least inconclusive, then there is no good reason to believe free will scepticism. This isn't a proof that some persons have free will but in a world of philosophical uncertainty it might be the best that we can do (Campbell 2011, 5.4-5).

Lastly, it is arguable that arguments for incompatibilism depend on *closure principles* of some kind ((van Inwagen 1983, pp. 56-57); (van Inwagen 1994, pp.95-98)). These principles are historically important and broadly similar to principles used in other sceptical arguments, especially knowledge closure principles used in the argument for epistemological scepticism. If we accept that incompatibilism is an essential premise in the argument for free will scepticism, then perhaps all sceptical arguments have underlying formal similarities. In this case, analogous responses might be given to them all (Strawson (1985); (Campbell 2011, 5.4)). That might be asking too much. Nonetheless, the study of arguments for incompatibilism offers an opportunity for free will scholars to connect with a broader set of issues, especially in epistemology, philosophical logic, and the history of philosophy.

In the last three decades the discussions on the so-called "Consequence Argument" have convinced many philosophers that compatibilism is not a viable theoretical option. What is your opinion on that argument?

The consequence argument is a brilliant and compelling argument for incompatibilism. I have argued that the consequence argument fails to prove incompatibilism (Campbell (2007); Campbell (2010)) but that does not detract from its brilliance and importance. This is especially true of several formal versions of the argument given over the last half century (Fischer (1994); Ginet (1966); Ginet (1990); van Inwagen (1983)).

The *consequence argument* assumes that there is a *deep past*: a past prior to the existence of human beings. It uses the claim that no one has or ever had a choice about propositions about the deep past – for no one was around then to have a choice –, to *ground* our lack of choice about propositions about the future in a deterministic universe. Given *determinism*, propositions about the past together with the laws of nature entail propositions about the future.

Yet one cannot get to free will scepticism from grounding principles and determinism alone. One needs a *transfer principle*: something to transfer our lack of choice about the past and laws onto a lack of choice about the future. Thus, versions of the consequence argument depend on closure principles such as:

β' From Np, p entails q deduce Nq,

where "Np" means "p and no one has or ever had a choice about whether p." Given principle (β') and determinism, our lack of choice about propositions about the deep past transfers onto propositions about the future, and free will scepticism follows (van Inwagen 1983, pp. 93-104).

Keep in mind that incompatibilism – as I understand the thesis cf. (Vihvelin 2008) –, is the view that determinism entails free will skepticism. Thus, if it is possible that someone has free will *and* determinism is true, then compatibilism is true. Alternatively, if incompatibilism is true, then it is impossible for anyone to have free will given determinism. Usually versions of the consequence argument assume the *classical* view of free will: free will requires the ability to do otherwise, though I'm not certain this assumption is essential. The conclusion – "no one has or ever had a choice" – can be given a variety of interpretations yielding a variety of different arguments and compatibility problems.

My criticism is that whatever the consequence argument proves it does not prove classical incompatibilism. For the argument assumes a deep past and it is not a necessary truth that the past is deep. One could imagine Adam existing contemporaneously with God, like an eternal flame sustained daily by God's omnipotent power. Suppose also that determinism is true. Still, one cannot use the consequence argument to show that Adam lacks free will. Hence, the consequence argument is not an argument for incompatiblism, other laudable merits notwithstanding.

This is a complex issue. On the one hand, Andrew Bailey (2012) shows that my criticism of the consequence argument is extended to other arguments for incompatibilism. On the other hand, there are versions of the consequence argument that

do not make an explicit reference to the necessity of the past ((van Inwagen 1983, pp. 78-93); Ginet (1990); Fischer (1994); Bailey (2012)). It is harder to show that these arguments fall victim to my criticism. Such arguments appeal to principles like the following:

> *The extension principle*: "an agent can do X only if his doing X can be an extension of the actual past, holding the laws fixed" (Haji 2009, p. 88)

This strikes me as the conclusion of a classical version of the consequence argument, not a premise in such an argument. Not surprisingly, Fischer & Pendergraft (2012) have replied to the charge that the consequence argument begs the question.

I confess that the consequence argument still has some force. Often we judge that someone lacks free will or moral responsibility because we think her action is the inevitable result of some prior set of circumstances over which she had no control. To the extent that one's behaviour is the result of genetic predispositions or childhood abuse and trauma, for instance, we judge that the behaviour is not up to the agent. R. Jay Wallace suggests that closure principles like (β') are the result of generalizing over particular cases like the ones noted above (Wallace 1994, pp. 16-17; pp. 114-116). Wallace and I agree that this generalization strategy fails but the compatibilist owes an explanation for why the particular inferences seem to work if not for faulty generalization.

Assuming that libertarianism as such is a viable position, which of the possible libertarian views (such as those centered on agent causation, indeterminist causation or no causation at all) are preferable?

All libertarian views are equally problematic (Campbell 2011, 3.3). Suppose a woman is choosing between two men to marry: Barry or Robin. She has equally good reasons for choosing either. As it happens, she chooses Barry. Suppose also she gets to make that same choice over again 99 more times, each time in the exact same set of circumstances: her reasons are the same, her memories are the same, the worldly facts are the same, the laws of nature are the same, etc. Every true proposition leading up to her action is held fixed yet she gets to make the same choice, in the exact same situation, 100 times over.

If it turns out that 100 out of 100 times she choses to marry Barry, then her choice seems inevitable, compelled. What kind of choice does one have if one always makes the same choice? Suppose that 50 times she chooses Barry and that 50 times she chooses Robin. How does this prove that she has free will? In this case, whether she chooses Barry or Robin is a matter of luck (2011, 1.5).

Libertarians try to get around this problem by providing detailed analyses of free action and compatibilists try to do the same. In both cases, the fundamental problems persist. This is why, though I have a compatibilist story to tell, I don't

dwell on it. Ultimately the problem is too confounding for any particular solution.

During the last years, a growing number of philosophers and scientists have advocated sceptical, eliminativist, pessimistic, or illusionistic views on free will. What do you think of these kinds of views?

Sceptical arguments play an important role in philosophical debates, so the new brands of free will scepticism have much to offer. Pereboom (2001) shows that we can keep a lot of moral theory even if philosophical scepticism turns out to be true. Smilansky (2001) has even argued that among the things that we can also keep is the *belief* that we have free will, agreeing with P.F. Strawson (1985) that sceptical arguments do little to dislodge our ordinary beliefs. Thus, free will scepticism might not be as pessimistic as it sounds.

It is possible that there is a mere verbal dispute between free will scepticism and compatibilism. The free will sceptic has a very high standard for what counts as "free." According to Galen Strawson, for instance, the kind of freedom relevant to moral responsibility requires that we be the ultimate source of our actions. Strawson further argues that this is impossible whether or not determinism is true. One might reply that ultimate sourcehood is too strict a criterion for either free will or moral responsibility. After all, who could be the ultimate source of *anything* other than God? Finding out that we are not gods is not worrisome, not as worrisome as finding out that we lack free will or that no one is morally responsible for anything. There appears to be a linguistic bait-and-switch.

Again, the comparison with epistemological scepticism is helpful. There was a time when philosophers were *infallibilists* about knowledge and believed that if one was justified in believing that a proposition was true, then the proposition must be true; that a truly justified belief was incapable of error. Sometime between Descartes' *Meditations* – where the clarity and distinctness of a judgment were endorsed as criteria for its infallibility – and Edmund Gettier's famous paper (Gettier 1963), we became *fallibilists* and decided to accept that knowledge did not require infallible certainty; that it was possible for one to be completely justified in a belief yet still be wrong about it. We didn't lament over this change. We merely adopted a more realistic conception of *knowledge*. Perhaps we're still awaiting a similar transition for the concept of *free will*. Free will is not ultimate *sourcehood* – only God could have that. It is merely partial or *adequate sourcehood* (Campbell 2011, 4.3).

The above story makes it seem as if the term "sourcehood" is ambiguous but there might be a more complex semantic story to tell. Perhaps what the different sceptical arguments have in common is that they each exploit the vagueness of a philosophical word or concept. Even in the case of the concept of *knowledge* it isn't as if our infallibilist intuitions have completely gone away. You think you know

you have a hand but do you *really* know? After all, you could be a brain-in-a-vat! In a similar way, claims about our abilities might be questioned. In most contexts, if I say that I am able to ride a bike the utterance is true but the claim might be challenged if it happens that there is no bike available to ride.

A very recent debate concerns the nature of our pre-philosophical views regarding free will. However, some surveys seem to suggest that we naturally tend towards compatibilism, others that we naturally tend towards incompatibilism. What do you think is the value of this kind of "experimental philosophy" in regard to the issue of free will?

Experimental philosophy is an exciting new philosophical field. Philosophical arguments often make appeals to common sense in order to support their views. Some free will theorists, for instance, contend that incompatibilism is the intuitive view and that compatibilism is unintuitive. These are empirical claims that require empirical evidence. Here the importance of experimental philosophy is undeniable.

Unfortunately, as the question indicates, there is no consensus among experimental philosophers about what the folk think about these matters. Some believe that the folk are incompatibilists, some believe that they are compatibilists. It has even been said that the folk concept of *free will* is incoherent, containing both compatibilist and incompatibilist criteria.

Is "free will" a folk expression or a term of art? Did the term "knowledge" change meaning when philosophers transitioned from infallibilism to fallibilism? Or did we come to learn that knowledge didn't have the properties that we once thought it had? How, if at all, does philosophical debate influence the meaning of philosophical expressions? I'm not sure how to answer these questions and equally unsure that experimental philosophy is going to help us (cf. Vargas (2013)). It seems that the empirical data equally supports a variety of theories about the meaning of philosophical terms and expressions.

Compatibility problems have been around as long as people have been thinking about free will and moral responsibility. Our concept of *knowledge* appears to have undergone a transition from infallibilism to fallibilism but the concept of *free will* has failed to undergo a similar transition. We are still burdened by medieval misconceptions. Manuel Vargas (2013) would call this our folk intuitions.

Incompatibilism requires a substantive argument, like the consequence argument. If "free will" means "blah blah blah *and* determinism is false," then the argument for incompatibilism is just a piece of analytical reasoning. It is difficult to understand what philosophers have been debating about all these years. Apparently, many of them simply failed to track the meaning of certain common terms and expressions in their language.

That incompatibilism is true, if it is true, strikes me as a substantive conclusion. It isn't just a matter of cashing out the meaning of the expression "free will." On my view, the argument for incompatibilism is fallacious. People who think otherwise are making some mistake: Either one thinks a closure principle like (β') is valid yet it is invalid or one accepts some false principle about the necessity of the past or the laws.

What do you think the relationship is between free will and moral responsibility? With regard to this, do you think that the famous Frankfurt scenarios are crucial for assessing the issue?

There is a "free will" crisis, a crisis about the meaning of "free will," that resulted from the upheaval caused by Harry Frankfurt's paper (Frankfurt (1969); (Campbell 2011, 2.4)). Prior to Frankfurt's essay, most philosophers accepted these propositions:

   (a) The ability to do otherwise is necessary for free will.

   (b) Free will is necessary for moral responsibility.

(a) and (b) together entail the principle of alternative possibilities:

   PAP The ability to do otherwise is necessary for moral responsibility.

Frankfurt supposedly showed that PAP is false and philosophers who accept Frankfurt's conclusions have a choice: either continue to hold (a) and reject that moral responsibility requires free will or continue to hold (b) and reject that free will requires the ability to do otherwise. Of course, some of us reject that Frankfurt examples are genuine counterexamples to PAP (Campbell (1997); Campbell (2005)).

In my book, I tried to provide an account of "free will" that was amenable to all parties in these debates. A person has *free will* provided that some of her actions are up to her. *Classical free will* theorists hold in addition to this that an act was up to an agent only if she was able to do otherwise (1.2; 2.4). The debate, then, is whether or not the relevant up-to-us ness requires the ability to do otherwise. This would end the crisis without settling any substantive debates.

For the record, I still think Frankfurt examples fail as counterexamples to PAP for precisely the same reasons that I hold classical compatibilism. I'm split between two replies: (i) terms like "able" are ambiguous between an irrelevant sense – perhaps even an incoherent sense –, that is incompatible with indeterminism and a relevant sense that is not, and (ii) the argument for incompatibilism suffers from a deeper linguistic malady than the fallacy of ambiguity, likely a contextualist fallacy of some kind.

Given the evidence coming from neuroscience and genetics, during the last few years a growing number of scholars have been arguing that the idea that we deserve blame for our bad deeds (and punishment for the worst of them) is ungrounded and should be abandoned. What is your opinion of this view?

There is no denying that the empirical evidence from neuroscience is compelling and will ultimately lead to a better understanding of free will and moral responsibility. Most free will scholars are convinced that the threats from neuroscience and genetics provide a new set of problems for free will, different than compatibility problems. Not surprisingly, I disagree.

One compelling set of data come from experiments performed by Benjamin Libet and others (Bayne (2011); Haggard (2011); (Swinburne 2011, introduction)). In a *Libet experiment*, a subject is asked to perform some simple motor action, like flexing her wrist, at some particular time of her choosing. The subject is then requested to identify the time at which she is first consciously aware of her decision (or intention) to flex her wrist. Studies reveal preparatory brain activity called the "readiness potential": a neural event or set of events that precede conscious choice yet is highly correlated with a particular action. Given common sense assumptions, free will appears to be an illusion, for decisions are fixed prior to conscious choice; conscious choice appears to play no role in one's decisions.

There is much disagreement about what the Libet experiments actually prove and there are several criticisms that one might give to any argument for free will scepticism based on these findings (Bayne 2011). What interests me is whether these findings present a new challenge to free will, something different from more traditional challenges resulting from various compatibility problems. Most philosophers believe that conscious choice has a neurological basis. Thus, there is likely some physical event or set of events in the causal chain leading from the readiness potential to the action that is token identical to the agent's conscious choice. The view is that this choice is nonetheless *bypassed* and plays *no causal role*, so it is no choice at all. Yet why believe that the physical event that is the agent's conscious choice plays no causal role in the agent's action? I have a hard time thinking that a no-choice closure principle of some kind is not needed to transfer our lack of choice about whether the readiness potential occurs to a lack of choice about whether the action occurs. On this telling of the story, it seems like the "new" challenge to free will is not so new after all.

This is why I adopt the compatibilist stance, trying to maintain compatibilism against all reasonable challenges. It is not that I am convinced that compatibilism is true. After all, what could prove a compatibility thesis? It is just that most people *do* have incompatibilist intuitions. Incompatibilist intuitions are so prevalent, that skilled philosophers often make use of essential features of the consequence argument without even noticing it.

Further, the similarities between various sceptical arguments hold hope for a common solution. Sure it seems that we have no free will given both empirical threats and abstract philosophical argument. It might also seem that nothing is flat and that we don't know that we have hands. You can easily render someone confused enough to believe that nothing is flat but you can't do this while simultaneously making her believe that the conclusion matters a great deal. I'm trying to promote a similar approach to free will.

Note

References to Campbell 2011 are given by chapter, section. Thus, "Campbell 2011, 5.4" means "Chapter 5, section 4." Standard references are given elsewhere.

References

Bailey, A. (2012), 'Incompatibilism and the past', *Philosophy and Phenomenological Research* **85**, 351–376.

Bayne, T. (2011), Libet and the case for free will skepticism, *in* R. Swinburne, ed., 'Free Will and Modern Science', Oxford University Press.

Campbell, J. (1997), 'A compatibilist theory of alternative possibilities', *Philosophical Studies* **88**, 319–330.

Campbell, J. (2005), 'Compatibilist alternatives', *Canadian Journal of Philosophy* **35**, 387–406.

Campbell, J. (2007), 'Free will and the necessity of the past', *Analysis* **67**, 105–111.

Campbell, J. (2010), 'Incompatibilism and fatalism: Reply to loss', *Analysis* **70**, 71–76.

Campbell, J. (2011), *Free Will*, Polity Press.

Fischer, J. M. (1994), *The Metaphysics of Free Will: An Essay on Control*, Blackwell Publishers.

Fischer, J. M. & Pendergraft, G. (2012), 'Does the consequence argument beg the question?', *Philosophical Studies* .

Frankfurt, H. (1969), 'Alternate possibilities and moral responsibility', *The Journal of Philosophy* **66**, 828–839.

Gettier, E. (1963), 'Is justified true belief knowledge?', *Analysis* **23**, 121–123.

Ginet, C. (1966), Might we have no choice?, *in* K. Lehrer, ed., 'In Freedom and Determinism', Random House.

Ginet, C. (1990), *On Action*, Cambridge University Press.

Haggard, P. (2011), Does brain science change our view of free will?, *in* R. Swinburne, ed., 'Free Will and Modern Science', Oxford University Press.

Haji, I. (2009), *Incompatibilism's Allure: Principle Arguments for Incompatibilism*, Broadview Press.

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

Smilansky, S. (2001), 'Free will: From nature to illusion', *Proceedings of the Aristotelian Society* **101**, 71–95.

Strawson, G. (2010), *Freedom and Belief, Revised Edition*, Oxford University Press.

Strawson, G. (2011), The impossibility of ultimate responsibility, *in* R. Swinburne, ed., 'Free Will and Modern Science', Oxford University Press.

Strawson, P. (1985), *Skepticism and Naturalism: Some Varieties*, Methuen.

Swinburne, R. (2011), *Free Will and Modern Science*, Oxford University Press.

van Inwagen, P. (1983), *An Essay on Free Will*, Clarendon Press.

van Inwagen, P. (1994), 'When the will is not free', *Philosophical Studies* **75**, 95–113.

Vargas, M. (2013), *Building Better Beings*, Oxford University Press.

Vihvelin, K. (2008), Compatibilism, incompatibilism, and impossibilism, *in* J. Hawthorne, T. Sider & D. Zimmerman, eds, 'Contemporary Debates in Metaphysics', Blackwell Publishing.

Wallace, R. (1994), *Responsibility and the Moral Sentiments*, Harvard University Press.