The Philosophical Quarterly
The Scots Philosophical Association and the University of St Andrews
The Philosophical Quarterly Vol. 63, No. 250 January 2013
ISSN 0031-8094 doi: 10.1111/j.1467-9213.2012.00094.x

FREE ACTS AND CHANCE: WHY THE ROLLBACK ARGUMENT FAILS

By Lara Buchak

The 'rollback argument,' pioneered by Peter van Inwagen, purports to show that indeterminism in any form is incompatible with free will. The argument has two major premises: the first claims that certain facts about chances obtain in a certain kind of hypothetical situation, and the second that these facts entail that some actual act is not free. Since the publication of the rollback argument, the second claim has been vehemently debated, but everyone seems to have taken the first claim for granted. Nevertheless, the first claim is totally unjustified. Even if we accept the second claim, therefore, the argument gives us no reason to think that free will and indeterminism are incompatible. Furthermore, seeing where the rollback argument goes wrong illuminates how a certain kind of incompatibilist, the 'chance-incompatibilist,' ought to think about free will and chance, and points to a possibility for free will that has remained largely unexplored.

Libertarians hold that free will is incompatible with determinism, but that we nonetheless have free will. Of course, the truth of indeterminism is not enough to guarantee free will: for an act to be free, it must originate from the agent herself in some important sense. Whether an act is free thus depends on the source of the indeterminism. We might take for granted that there are sources of indeterminism conducive to free acts. Recently, however, Peter van Inwagen has introduced an argument that has come to be known as the 'rollback argument,' that challenges whether indeterminism in any form can leave room for freedom. This argument purports to show that if indeterminism holds, then regardless of what this indeterminism consists in, *every* act is a mere matter of chance in the sense incompatible with free will. If the rollback argument is sound, then libertarians must conclude that free will is compatible with neither determinism nor its denial, and so, in the words of van Inwagen, 'free will remains a mystery.'

Determinism is the thesis that the state of the world at time t_1 in conjunction with the laws of physics entail the state of the world at a later time t_2. Libertarians hold that determinism is incompatible with free will, usually on the grounds that if there is only one physically possible future,

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

then an agent's actions are not 'up to' him in the sense relevant for free will. Indeterminism is just the denial of determinism, though it is clear that not just any kind of indeterminism will do for free will. For example, if an agent's actions at t_2 are undetermined at t_1 because they are to be determined by the flip of a coin between t_1 and t_2, then the agent's actions are not up to her any more than if they are determined at t_1. They are *mere matters of chance*. Free acts, according to libertarians, need to be not only undetermined, but undetermined in the right way: undetermined because they are ultimately up to the agent.

Van Inwagen's rollback argument challenges the idea that acts can ever be undetermined in the sense required for free will. The argument purports to show that regardless of what governs agent acts under indeterminism, all agent acts will have the same status as acts governed by coin-flips, which is to say, they will not be free. Van Inwagen specifically argues that agent-causation is not sufficient to make agent acts free, but his argument easily generalises to any way of spelling out what holds of an agent in an indeterministic world.

Here is his argument.[1] Consider an agent, Alice, who is deciding whether to lie. Let us assume her choice is undetermined by the state of the world at t_1 and the laws of physics. And let us say she lies at t_2. Can this have been a free act? To show that it cannot have been, van Inwagen asks us to consider what would have to be true if, hypothetically, God were to reset the universe to t_1 and let events transpire as they may; and if God were to do this many times over. Since Alice's lying is not determined, it would have to be the case that she would lie in some replays and not lie in others. Now, if God were to replay the event enough times, the proportion of replays in which Alice lies to replays in which she tells the truth would almost certainly converge to some definite number. For example, let's say that after 100 replays, she has lied 35 times; after 1000 replays, she has lied 326 times, and after 10000 replays she has lied 3076 times. We would then be confident that the proportion of lies to total cases would settle out to 0.3: she lies in 30% of the cases. But to say that she lies in 30% of the cases is just to say that there is a 30% chance of her lying in any particular case, including some hypothetical next case. And including, indeed, the actual case at hand. Furthermore, if there is a definite objective probability to her lying, then whether she lies in the case at hand is a mere matter of chance: it is as if whether she lies is determined by the flip of a biased coin which has a 30%

[1] P. van Inwagen, 'Free Will Remains a Mystery', *Philosophical Perspectives*, 14 (2000), pp. 1–20, at pp. 13–18.

chance of landing heads. Finally, notice that to reach this conclusion we did not rely on a particular assumption about the source of the indeterminism or the source of its resolution between t_1 and t_2: regardless of the mechanics of choice, says the argument, an undetermined choice is *relevantly like* flipping a coin.

To see the crucial steps of the argument, here it is in premise-conclusion form:

(P1) If indeterminism holds, then if God replayed the universe numerous times in the above scenario, it would become increasingly likely, as the number of replays increased, that the ratio of lies to truths would converge to some definite real number.

(P2) If the ratio of lies to truths would converge to a definite real number in the above scenario, then Alice's lying in the case at hand and Alice's telling the truth in the case at hand each have a definite objective probability at t_1, namely the ratio of lies to total cases and the ratio of truths to total cases.

(C1) If indeterminism holds, then Alice's lying and Alice's telling the truth each have a definite objective probability at t_1.

(P3) If an act has a definite objective probability at a time, then it cannot be a free act at that time.

(C2) If indeterminism holds, then whatever Alice does, it won't be a free act.

Discussion of the rollback argument has centered around (P3): denials of this claim are articulated by Mark Balaguer, Michael Almeida and Mark Bernstein, Timothy O'Connor, Laura Eckstrom, and Christopher Evans Franklin; and Seth Shabo provides an additional argument in its favour.[2] However, to my knowledge, everyone who discusses the argument has taken (P1) and (P2) for granted.[3] Denying (P2) is not a

[2] M. Balaguer, Free Will as an Open Scientific Problem, (MIT Press, 2010), chapter 3; M. Almeida and M. Bernstein, 'Rollbacks, Endorsements, and Indeterminism', in R. Kane (ed.), The *Oxford Handbook of Free Will*, (Oxford UP, 2011), pp. 484–95; T. O'Connor, 'Agent-Causal Theories of Freedom', in R. Kane (ed.), *The Oxford Handbook of Free Will*, (Oxford UP, 2011), pp. 309–28; L. Eckstrom, 'Free Will, Chance, and Mystery', *Philosophical Studies*, 113 (2003), pp. 153–80; L. Eckstrom, 'Free Will Is Not a Mystery', in R. Kane (ed.), *The Oxford Handbook of Free Will*, (Oxford UP, 2011), pp. 366–80; C. Evans Franklin, 'Farewell to the luck (and *Mind*) argument', *Philosophical Studies*, 156 (2011), pp. 199–230; S. Shabo, 'Why Free Will Remains a Mystery', *Pacific Philosophical Quarterly*, 92 (2011), pp. 105–25.

[3] In their discussion of the rollback argument, Almeida and Bernstein (pp. 485) and Franklin (pp. 216) explicitly endorse (C1) without argument, so we may assume they endorse (P1) and (P2). Eckstrom does note in passing that if an agent's choices are ungoverned by laws, they will have no probability (a denial of (C1)), though she doesn't detail where the rollback argument goes wrong if this holds or what 'ungoverned by laws' means in this context, since her focus is a denial of (P3).

very attractive option, since it seems to be an unproblematic instance of inference to the best explanation. But I claim that we have no reason to accept (P1).

It is worth looking in detail at why van Inwagen thinks that the ratio of lies will converge to some definite real number. Here is what he says:

> 'Now let us suppose that God *a thousand times* caused the universe to revert to exactly the state it was in at t_1 (and let us suppose that we are somehow suitably placed, metaphysically speaking, to observe the whole sequence of "replays"). What would have happened? What should we expect to observe? Well, again, we can't say what would have happened, but we can say what would *probably* have happened: sometimes Alice would have lied and sometimes she would have told the truth. As the number of "replays" increases, we observers shall—almost certainly—observe the ratio of the outcome "truth" to the outcome "lie" settling down to, converging on, some value…"Almost certainly" because it is *possible* that the ratio not converge. Possible but most unlikely: as the number of replays increases, the probability of "no convergence" tends to 0.'[4]

Van Inwagen's reason for thinking that the convergence will occur is clearly the law of large numbers, which says roughly that if we repeat an event with two possible outcomes many times over, the ratio of each outcome to the number of trials will, with increasing likelihood, tend to the (objective) probability of each outcome. For example, if we flip a biased coin long enough, the proportion of heads to total flips will almost certainly converge to the coin's bias towards heads.

However, van Inwagen fails to notice that there is an important difference between the coin case and Alice's case. In the case of the coin, we apply the law of large numbers because we assume the coin *does* have some definite objective probability of landing heads. That there is some definite probability involved is a *presupposition* of the law of large numbers. For example, here is a typical statement of the law:

> 'In repeated, independent trials with the same probability p of success in each trial, the percentage of successes is increasingly likely to be close to the chance of success as the number of trials increases. More precisely, the chance that the percentage of successes differs from the probability p by more than a fixed positive amount,

[4] Van Inwagen (p. 14 and footnote 16). The passage I have quoted is from van Inwagen's argument that undetermined acts aren't free, without allowing for the possibility of agent-causation. He goes on to claim that the argument works the same way for agent-caused undetermined acts, since it is nowhere mentioned whether or not Alice's acts result from agent-causation. We may also assume that it is supposed to work against any kind of act under indeterminism, since it nowhere relies on the mechanics of action or of indeterminism.

$e > 0$, converges to zero as the number of trials n goes to infinity, for every number $e > 0$.'[5]

We can only apply the law at all if its antecedent is satisfied: i.e., if the event in question has some probability p, and has this probability in each of the trials. But this is precisely what van Inwagen is trying to argue *for* in this step of the argument: he is trying to argue that we can assign a probability to the event that Alice lies.[6]

The rollback argument directly begs the question of whether Alice's lying has an objective probability. Without the assumption that it does, there is nothing at all in the setup of the rollback scenario itself to guarantee the truth of (P1). There is nothing at all to rule out, for example, the following series of choices: the first time God reruns the situation, Alice lies; the next 9 times, she tells the truth; the next 90 times, she lies; the next 900 times, she tells the truth; and so forth. In this example, the proportion of lies never converges (it will alternate between roughly 1/11 and 10/11, after each 10^n trials). Contra van Inwagen, *there is nothing in his setup even to make this unlikely*. Unlike in the coin-flipping case, there may not be a chancy mechanism – or a mechanism that *behaves as if* it is governed by chance – grounding Alice's actions. Since (P1) and (P2) are supposed to supply an argument for (C1), van Inwagen can't support (P1) using the law of large numbers, because to do so assumes (C1), the very thing at issue. The truth of (P1) is an empirical question, and one we are incapable of testing in principle.

It is now clear that we have no reason to be convinced by the rollback argument as it stands. But the insight here goes beyond a refutation of the rollback argument. That one can accept (P3) without concluding indeterminism and free will are incompatible points to an unexplored possibil-

[5] P.B. Stark, 'Glossary of Statistical Terms,' http://www.stat.berkeley.edu/~stark/SticiGui/Text/gloss.htm. Accessed on 2/09/2011.

[6] After establishing that the replays would converge to a definite proportion, van Inwagen imagines for illustration that the proportion of lies to truths is roughly even, and then writes (p. 15): 'A sheaf of possible futures (possible in the sense of being consistent with the laws) leads 'away' from $[t_1]$, and, if the sheaf is assigned a measure of 1, surely, we must assign a measure of 0.5 to the largest sub-sheaf in all of whose members Alice tells the truth and the same measure to the largest sub-sheaf in all of whose members she lies. We must make this assignment because it is the only reasonable explanation of the observed approximate equality of the 'truth' and 'lie' outcomes in the series of replays. And if we accept this general conclusion, what other conclusion can we accept about the seven-hundred-and-twenty-seventh replay (which is about to commence) than this: each of the two possible outcomes of this replay has an objective, 'ground-floor' probability of 0.5—and there's nothing more to be said? And this, surely, means that, in the strictest sense imaginable, the outcome of the replay will be a matter of chance.' Thus, it is clear that the rollback scenario is supposed to *establish* that lying has a definite probability, namely a probability equal to the (convergent) proportion of cases in which the agent lies.

ity for 'chance-incompatibilists,' i.e., incompatibilists who think that an act cannot have been free at a time if its occurrence had a definite chance at that time. In particular, it is open to chance-incompatibilists to deny that a free act has a definite objective chance of occurring before the agent exercises her free will.

The thought that there is a difference between agent acts and ordinary goings-on in the world – in this case a difference in whether we can assign objective probabilities to their occurrence ahead of time – naturally calls to mind the original target of van Inwagen's argument: agent-causation. Agent-causation views say that an act is free just in case the agent in question is a 'substance' that acts rather than a mere locus for physical events in the causal chain of that act: this is to say, if we list only the physical events leading up to a free act, then we have left out a member of the causal chain.[7] The metaphysics of agent-causation are notoriously tricky, but the discussion here points us to one concrete metaphysical difference that the proponent of agent causation could postulate: agent-caused events lack objective probabilities. Of course, introducing agent-causation is not the only way for the chance-incompatibilist to deny that agent acts have objective probabilities. There may be other theories about the metaphysics of free will that can plausibly deny this. The point is that there are avenues open to the chance-incompatibilist to resist the conclusion that we lack free will: as long as one analyses free will in such a way that free acts lack objective probability, van Inwagen's argument will have no purchase.

Is maintaining that free acts lack objective probability inconsistent with what current physics tells us? While a full discussion of this question goes beyond my knowledge of physics, here is a reason to think that it is not. This reason originates in an argument for a seemingly unrelated point: specifically, in Alan Hájek's argument that conditional probability rather than unconditional probability ought to be thought of as primitive.[8] In the course of his argument, he notes that quantum mechanics primarily tells us about certain objective *conditional* probabilities. For example, he says that the 'Born rule' tells us about probabilities of the form $p(O_k \mid M)$, where M is the proposition that a particular measurement takes place (according to Hájek, the act of some agent) and O_k is the proposition that

[7] The classic statement of this view can be found in R.M. Chisholm, 'Human Freedom and the Self', University of Kansas Lindley Lecture, Department of Philosophy, University of Kansas (1964), pp. 3–15. Reprinted in G. Watson (ed.), *Free Will*, (Oxford UP, 2003), pp. 26–37.

[8] A. Hájek, 'What Conditional Probability Could Not Be', *Synthese*, 137 (2003), pp. 273–323.

a particular outcome eventuates.[9] Hájek argues that quantum mechanics itself (QM uninterpreted) does not assign an unconditional probability to the proposition M: it is silent on $p(M)$. He argues further that quantum mechanics cannot in principle deliver probabilities of the form $p(M)$. I will not rehash his arguments here. And while Hájek's conclusion is not uncontroversial (and he states as much),[10] the point for present purposes is that physics hasn't made up its mind about whether all events – in particular events involving the actions of agents – have objective unconditional probability. Indeed, Hájek cautions us against inferring from the fact that the micro-level events which are the central subject of physics have probabilities relative to the measurements of observers to the claim that all events have unconditional probabilities:

> 'It seems to me that the intuition that chances must always exist, even for free acts, parallels the intuition that values for observables (such as position and momentum) must always exist. But the latter intuition has been challenged since Bohr, and has hit particularly hard times since the Kochen-Specker theorem.' (307)

We shouldn't be too quick to assume that our current physical theories will assign objective chance to acts, nor that they will say the same things about the behaviour of agents that they do about the behaviour of particles. They might or might not, but it is an empirical question we are not currently in a position to answer.

If Hájek's argument is right, then physics is not committed to assigning unconditional chances to free acts – and there may be additional reasons to think that we cannot assign them. However, we typically will be able to assign *conditional* chances to propositions. So the important question will be what sorts of conditional chances we can assign to agent acts at the time when they are purportedly free, and whether being able to assign these is incompatible with free will. For example, we might ask which conditional chances of Alice lying at t_2 get assignments at t_1: if A is the proposition that Alice lies at t_2, for what conditions $\{C\}$ does $p(A \mid C)$ have

[9] Interestingly enough, earlier in the article and in quite a different context than the discussion in this paper, Hájek (p. 304) uses what is essentially a modus tollens version of van Inwagen's argument to show that not all propositions have unconditional relative frequency (relative frequency being a candidate interpretation for objective probability). Hájek asks us to consider an agent freely deciding whether to toss a coin, and points out that the agent could decide to deliberately make choices so that the frequency with which he decides in the affirmative fluctuates wildly over time, i.e., so that the sequence has no limiting frequency. Of course, this isn't an argument that we do have free will (and Hájek certainly doesn't intend it to be!), but his use of the modus tollens argument further shows that the truth of (P1) ought not be considered settled in contexts outside of this debate.

[10] Hájek (p. 307) notes that Bohm's interpretation, collapse interpretations, and the many worlds interpretation all imply that probabilities of the form $p(M)$ are well-defined.

a determinate value at t_1? And we can then ask whether the existence of any of these conditional chances ought to worry us.

Simply showing that there are some conditional chances of the form $p(A \mid C)$ won't reveal a problem for the claim that A is a free act. If C is merely a physical description of Alice's lying at t_2, then $p(A \mid C)$ will equal 1, but this is surely not troublesome. For in this case, the physical description merely *is* the free act, and since $p(A)$ is not determinate, $p(C)$ is not determinate. More generally, if C is a description of some act whose objective unconditional probability is determinate, then the indeterminateness of $p(A)$ implies that at least one of $p(A \mid C)$ and $p(A \mid \sim C)$ is indeterminate. This is to say, conditional on at least one of C or $\sim C$ the act does not have a determinate objective probability – which I take it is all the chance-incompatibilist needs. So we should expect not to be able to pick a C such that $p(C)$, $p(A \mid C)$, and $p(A \mid \sim C)$ are all determinate at t_1. Therefore, conditional chances of the form $p(A \mid C)$ where C is some physical event that has a determinate objective probability should not ordinarily pose a problem for the claim that A can be a free act.

The chance-incompatibilist cannot, however, conclude that there won't be *any* conditional chances that will undermine freedom. For the lack of a determinate $p(A \mid C)$ for determinate $p(C)$ does not imply that Alice *does* have free will: if Alice's lying is determined by the free act of some other agent (if C is 'Mary forces Alice to lie'), it is surely not free. This observation draws attention to the fact that there are two ways in which probabilistic facts can entail that an agent-act A is not free, according to the chance-incompatibilist. The first is if $p(A)$ is determinate, which we already saw is not compelled by current physics (at least on some still-open interpretations). The second is if there is a determinate $p(A \mid C)$ where C is the free act of some other agent. If, at t_1, there is some definite probability of Alice lying conditional on an act of Mary's, chance-incompatibilists will presumably think that Alice is not free at t_1, or at least won't be free if Mary does perform the act: conditional on what Mary does, it is a mere matter of chance whether Alice will lie.

It is open to all chance-incompatibilists – proponents of agent-causation and otherwise – to deny that agent-acts have determinate probabilities. However, the second way in which probabilistic facts can threaten freedom sheds light on which kinds of chance-incompatibilists can claim that there are free acts without departing too radically from current physics. Current physics says that many conditional probabilities of the form $p(B \mid C)$ do exist: namely, conditional probabilities where C is the proposition that a particular measurement takes place and B is a description of a micro-level event of the type studied by physics. Therefore, if free acts are

just micro-level events of the type studied by physics, then there should be a determinate $p(A \mid C)$ where, for example, A is the proposition that Alice lies and C is a proposition describing some measurement process. Given this, the chance-incompatibilist has two ways to make room for freedom. First, she can deny that $p(A \mid \sim C)$ is determinate, and argue that whether an act is free depends on whether the agent is part of a system for which a measurement is in fact not taken; but to take this route she will have to spell out why not taking a measurement should make a difference to freedom. Second, she can deny that free acts are micro-level events of the type studied by physics. This is what the proponent of agent-causation denies. There may be other ways to deny this, but denying this without departing too radically from current physics depends on finding some way to distinguish between agent acts and other kinds of events such that the objective probabilities conditional on measurements won't always be determinate for agent acts even though they are for micro-events that don't involve agents. And this may be a difficult task for the theorist who thinks that the decisions of free agents have ordinary micro-level descriptions.

I have shown that libertarian freedom is not in as bad a spot as we might have thought. In particular, the rollback argument does not show, even for chance-incompatibilists, that free will is incompatible with indeterminism. If chance-incompatibilism is true, then the question of whether free will is compatible with determinism depends on what exactly agent acts are, and on what our best physical theory ultimately says about whether agent acts have objective chance. The discussion here points the way forward in two respects. First, it reminds us that taking physics seriously may be consistent with thinking there really is a difference between events involving free acts and other kinds of events. Second, it suggests that we ought to turn our attention to the question of what physics is actually committed to as regards the objective chances of acts involving agents, and whether what physics is committed to in this regard is incompatible with free will.[11]

University of California at Berkeley

[11] I would like to thank Mark Balaguer, Alan Hájek, Jeff Russell, and Neal Tognazzini, for helpful discussions and for their comments on earlier versions of this paper.