Two Intuitions about Free Will: Alternative Possibilities and Endorsement

Christian List and Wlodek Rabinowicz[1]

16 December 2011

1. Introduction

Many of us have the following two intuitions about free will:[2]

> **Intuition 1: "Free will requires the ability to do otherwise."** An agent's action counts as free *only if* the agent could have acted otherwise.
>
> **Intuition 2: "Free will requires more than making un-determined 'fluke' choices."** An agent's action counts as free *only if* the action is endorsed by the agent in an appropriate way, as opposed to having been merely indeterministically picked from some set of alternative possibilities, for instance by randomization or some contingency outside the agent's control.

For present purposes, we adopt a modal interpretation of the ability to act otherwise, according to which an agent can act otherwise if and only if it is *possible* for him or her to do so, in an appropriate sense of possibility.[3] According to Intuition 1, under this modal interpretation, a necessary condition for free will is indeterminism at the agential level.

> **Indeterminism at the agential level:** There are some points in time at which the agent's prior history has more than one possible future continuation, corresponding to different courses of action.[4]

Intuition 2, by contrast, suggests a rather different picture of free will, according to which free will is tied not to indeterminism at the agential level, but to the agent's endorsement of his or her actions. The juxtaposition of Intuition 1 and Intuition 2 now leads to the following worry.

> **The worry:** Suppose we have indeterminism at the agential level, in accordance with Intuition 1. The agent's action at time *t* is one of several equally possible actions, given his or her prior history up to *t*. In particular, the modal relationship between the actual action and the prior history is no different from the modal relationship in which any of

---

[1] C. List, Departments of Government and Philosophy, London School of Economics; W. Rabinowicz, Department of Philosophy, Lund University. We gratefully acknowledge the hospitality of the Swedish Collegium for Advanced Study, where we wrote this paper. We are also grateful to John Broome for helpful discussion.

[2] For surveys of the debate on free will, see Kane (2002) and Fischer, Kane, Pereboom and Vargas (2007).

[3] A defence of the modal interpretation of the ability to do otherwise is beyond the scope of this paper; see List (2011). We further leave open the question of what precise sense of possibility is required. Our analysis is consistent both with a metaphysical notion of possibility, which free-will libertarians may adopt, and with an agential (more coarse-grained) notion of possibility, which some free-will compatibilists may favour, especially if they interpret free will as an emergent, higher-level phenomenon (e.g., Dennett 2003, List 2011). The idea of agential possibility has also received attention under the label of "agentive modalities". See, e.g., Maier (2011) for an analysis of "agentive modalities" on the basis of the concept of an agent's options.

[4] Like the possibility of doing otherwise, the notion of indeterminism at the agential level can be interpreted in different ways, depending on how we spell out the notions of "possibility" and an agent's "history". What matters is that we use the same modal notions throughout the present paper.

> the other possible actions would have stood to that history. Each of these possible actions could have been actualized, and none of them – including the actual one – was specially supported, let alone necessitated, by the agent's prior history. It thus seems hard to attribute the actual action to anything other than a "fluke" or some impersonal, indeterministic process. According to Intuition 2, then, the agent's action cannot count as free.

In short, for an action to count as free, according to Intuition 1, it must *not* have been predetermined by the agent's prior history, whereas, according to Intuition 2, it must stand in some privileged relationship to that history: it must not have been just a "fluke". Can there be any free actions under these constraints?

The apparent tension between Intuitions 1 and 2 is one of the motivations behind the "hard incompatibilist" view, according to which free will in the conventional sense is compatible neither with determinism nor with indeterminism.[5] We aim to show, however, that the tension is only apparent, and that there *can* be free actions in accordance with both intuitions.

The key idea underlying our reconciliation of the two intuitions is that there is an important distinction between (i) which actions an agent can *possibly* do, and (ii) which actions he or she can do *with endorsement* (or "rationally do", for those who prefer the latter terminology). With this distinction in place, it is entirely consistent to say that an agent who made a particular choice had the ability to do otherwise – in the sense that several actions were possible for him or her, none of which was predetermined (in line with indeterminism at the agential level) – and yet that, far from having made a fluke choice, the agent chose an action that he or she endorsed.

Although Intuitions 1 and 2 can be reconciled in this manner, their reconciliation leads to an interesting and perhaps surprising qualification to our picture of free will: while we can consistently hold that free will requires the ability to act otherwise even if we accept Intuition 2, we cannot generally hold that it also requires the ability *freely* to act otherwise, where this is understood as the ability to act otherwise *with endorsement*.

2. A toy model

A toy model helps to present the architecture of our reconciliation of Intuitions 1 and 2.[6] We consider a world with a single agent and represent this world as a simple dynamic system. The system can be in a number of possible states, and its state evolves over time. We call the set of all possible states of the system its *state space* and label it $\mathbb{S}$. We further use T to denote the set of all points in time, where T is linearly ordered. It could be, for instance, the set of real numbers or the set of natural numbers; no assumption apart from linearity is needed.

An *agential history* is a temporal path of the system through the state space, formally a function, $\mathbb{h}$, from T into $\mathbb{S}$, which assigns to each point in time, t in T, a corresponding state of the system, $\mathbb{h}(t)$ in $\mathbb{S}$. We write Ω to denote the set of all agential histories that are possible relative to the laws

[5] See, e.g., Pereboom (2005).

[6] Some of the present formalism – specifically the modelling of the ability to do otherwise in terms of branching in agential histories – draws on List (2011). Cf. Segerberg (1984), Belnap, Perloff and Xu (2001) and Belnap (2005).

governing our agential system. As is standard, subsets of Ω can be interpreted as *propositions*, where a proposition is *true* in all histories contained in it and *false* in all other histories.

To express conditions such as determinism or indeterminism, we need to introduce the notion of *branching* in agential histories. It is helpful to use the notation h_t to denote the *truncated part* of history h up to time t. Formally, h_t is the restriction of the function h to the sub-domain of all points in time up to and including t. Note that h_t is itself a function, namely from the set of all points in time up to t into the set S. Importantly, the truncated history h_t must not be confused with $h(t)$, the agent's state at time t.

Now a *branching point* occurs in history h at time t if there exists another history h' which coincides with h up to time t and is distinct thereafter, formally, $h_t = h'_t$ and $h_{t'} \neq h'_{t'}$ for every point in time t' after t.[7] Our agential system is *deterministic* if none of the histories in Ω exhibits any branching point, and *indeterministic* otherwise. (Similarly, we can define determinism and indeterminism in relation to individual histories.)

We are now in a position to revisit our two intuitions about free will. Intuition 1 says that free choices require the ability to do otherwise. We interpret this as the claim that free choices can occur *at most* at branching points. Without branching points in agential histories, there cannot be free will. But we can also see why this indeterministic requirement on free choices may seem to conflict with Intuition 2. Any branching point is characterized precisely by the fact that the same truncated history up to time t can have two or more distinct continuations. Nothing in the agent's truncated history privileges one such continuation over any other – at least not if we refer only to the elements of the model introduced so far. And so there is nothing that could make the *actual* continuation of the agent's history, as opposed to its counterfactual alternatives, count as "specially endorsed" by the agent.

Just as a light particle with the same past trajectory can be indeterministically prompted to pass either through the left slit or through the right slit in the famous double-slit experiment, so an agent with the same past trajectory can be indeterministically prompted to continue either with one future trajectory or with another. Both continuations of the agent's history after time t stand in the same modal relationship (specifically: relationship of possibility) to the agent's history up to t.[8]

To capture the idea that only some but not all possible continuations of the agent's history at a given time count as "endorsed" by the agent, an extension of the present model is needed. We need to introduce one further primitive notion: an *endorsement function*. Formally, this is a function

- whose domain is a set of truncated histories (typically, histories that are truncated at branching points, but histories truncated at other points need not be ruled out by definition),
- and whose co-domain is a set of sets of histories,

[7] The fact that the *functions* $h_{t'}$ and $h'_{t'}$ differ for every point in time t' after t does not rule out the possibility that the *states* $h(t')$ and $h'(t')$ might coincide for some such t'.

[8] Note that we say all this against the background of indeterminism at the agential level. The modal notions used here refer to possibilities for the agent. As noted above, for the purposes of this paper, we do not take a view on how the notion of agential possibility relates to any underlying notion of metaphysical or physical possibility.

where

> any truncated history in the relevant domain is mapped to the set of those continuations of that truncated history which count as "endorsed" by the agent – or, in alternative terminology, as "rationally supported". (A *continuation* of a truncated history $\mathbb{h}_t$ is a history $\mathbb{h}'$ such that $\mathbb{h}'_t = \mathbb{h}_t$.)

Let us illustrate this notion. Suppose $\mathbb{h}_t$ is an agential history up to a particular point in time, such as a branching point at which the agent is faced with a decision. Provided that $\mathbb{h}_t$ is included in the domain of the endorsement function, $\mathbb{h}_t$ is then mapped by that function to the set of those continuations of $\mathbb{h}_t$ that the agent endorses (or that are rationally supported by the agent's state at time *t*). This will be a subset (not necessarily a proper one) of the set of all *possible* continuations of $\mathbb{h}_t$. We deliberately leave open the precise interpretation of "endorsement". This could be spelt out in a variety of ways, consistently with our formal framework.

Our definition of an endorsement function is also consistent with a number of different *formal* properties of endorsement. For example, an endorsement function could map some truncated histories to the empty set. This allows for the possibility that some truncated histories have no endorsed continuation. Similarly, an endorsement function could map some truncated histories to a non-singleton set of continuations. This allows for the possibility that some truncated histories have more than one endorsed continuation. By contrast, whenever a truncated history is mapped to a singleton set of continuations, there is a unique endorsed continuation of that history.

It is further helpful to distinguish between two different cases of a non-singleton set of endorsed continuations of a given history. It may be that

(i) several possible continuations of $\mathbb{h}_t$ count as endorsed, but
(ii) they all coincide up to some *later* branching point *t'* and diverge only thereafter.

In such a case, we say that the set of endorsed continuations of $\mathbb{h}_t$ is *effectively singleton* (though not literally singleton), since there is only one endorsed *immediate future* after time *t*, and up to time *t'*. By contrast, if (i) is true and (ii) is false (i.e., some histories in the set diverge from each other immediately after time *t*), we say that the set of endorsed continuations of $\mathbb{h}_t$ is *effectively non-singleton*.[9]

To ensure the generality of our framework, we do not impose any restrictions on the endorsement function. It should be clear that whenever the agent's truncated history at a particular branching point is mapped to a singleton or effectively singleton set of endorsed continuations, there is, in effect, only one immediate course of action that the agent can take *with endorsement*. The agent can still act otherwise at such a branching point, since there is more than one distinct future

[9] This complication has to do with a more general problem: it is plausible to say that the alternatives between which an agent in a history $\mathbb{h}$ potentially chooses at some time *t* need not be different fully specified histories that are continuations of $\mathbb{h}_t$ but can be different mutually exclusive *sets* of such continuations. Both our powers of discrimination and our capacities to control the future may be too limited to make the fully specific continuations available for choice. The case in which the set of endorsed continuations is effectively but not literally singleton illustrates this: what the agent endorses in this case is a single course of action that fixes the future up to a certain point, but leaves several possibilities open beyond that point.

continuation of the agent's history (this follows from the definition of a branching point), but only one branch counts as endorsed. It follows that:

> **Observation:** Although at any branching point the agent can act otherwise, he or she may not generally be able to act otherwise *with endorsement*. Taking an endorsed action at a given branching point is compatible with the ability to act otherwise with endorsement *if and only if* the set of endorsed continuations of the agent's history at that branching point is effectively non-singleton.

Crucially, the condition on the right-hand side of the last biconditional is not met in general.

3. The example of Martin Luther

An example that is frequently discussed in the literature on free will is that of Martin Luther.[10] When Luther was summoned to the Reichstag zu Worms in 1521 and was asked to renounce his critical views on the Roman Catholic Church, he reaffirmed his views, allegedly saying "here I stand; I can do no other". Does this mean Luther lacked free will on this matter? Most commentators agree that, far from renouncing free will, Luther was actually taking responsibility for his actions, implying that these were necessitated by his character and convictions.

Should we interpret this example as showing that free will does not require the ability to act otherwise, thereby refuting Intuition 1? Daniel Dennett suggests this interpretation.[11] We think, however, that this would be too quick. A better interpretation of Luther's assertion would be that although it was not literally impossible for him to act otherwise, he could not have acted otherwise *without sacrificing his integrity*.[12]

Let us recast this example in the terminology of the last section. When Luther arrived at the Reichstag zu Worms, he had in fact reached a branching point in his agential history. One continuation of this history would have involved a renouncement of his views; another – the one he actually pursued – a reaffirmation. Both continuations were literally possible, and what distinguished the second from the first was *not* its modal status *but* the fact that it, alone, was endorsed by Luther. Formally, Luther's endorsement function mapped his truncated history at the given time to a singleton (or effectively singleton) set: the set containing the history in which he reaffirmed his views. In short, Luther did have multiple possible futures at the time, but there was only one that he endorsed.[13]

These considerations suggest a distinction between two notions of a free action, which echoes a distinction familiar from related discussions, including Isaiah Berlin's discussion of freedom in a political, as opposed to metaphysical, context (1958). According to the first, "negative" notion, an

---

[10] See, e.g., Dennett (1984) and Kane (2002, introduction).
[11] See Dennett (1984).
[12] See also List (2011).
[13] For ease of exposition, we deliberately individuate histories in a very coarse-grained manner here, not differentiating between different possible versions of Luther's act of reaffirming his views (e.g., with respect to the precise choice of words, timing etc.). Luther himself invoked such a coarse-grained individuation of histories when he said "I can do no other". Formally, what matters is that Luther endorsed only those histories in which he reaffirmed his views, while there existed other possible histories that he did not endorse.

action counts as free if and only if the agent could have acted otherwise. This notion is sensitive only to Intuition 1. According to the second, "positive" notion, an action counts as free if and only if the action was endorsed by the agent. This notion is sensitive only to Intuition 2. Harry Frankfurt's famous argument that the ability to do otherwise is not necessary for moral responsibility may be described as relying on such a positive notion of freedom.[14]

Luther, we suggest, was free in both senses: he could have acted otherwise, and he did in fact endorse his action. But although he could have acted otherwise, his alternative was not one that he endorsed. If endorsement is necessary for freedom, then, although Luther could have acted otherwise, he could not have *freely* acted otherwise.

We do not think, however, that this last observation in any way challenges the claim that Luther acted out of his own free will in the given instance. While the exercise of free will may require *both* the ability to act otherwise *and* the endorsement of one's action (in accordance with Intuitions 1 and 2), it does not seem to require *the ability to act otherwise with endorsement*. The ability to act otherwise and the endorsement of one's action are arguably two separate dimensions of free will, which are not normally – and cannot generally be – entangled with one another.

4. An illustrative decision-theoretic interpretation of endorsement

We have been silent on how exactly to interpret the notion of endorsement (as well as the notion of agential possibility). Our aim has been to sketch a broad "architecture" for thinking about our two intuitions about free will, without adjudicating all the details, on which there may be considerable disagreement.

To show that a concretization of our formal framework can be given, however, we now provide one illustrative interpretation of its central concepts, drawing on standard decision theory.[15] This is not meant to be a definitive interpretation (indeed, we think that the standard decision-theoretic picture is in need of refinement[16]). Our aim is merely to show that our framework is naturally compatible with established theories of individual decision making.

In a standard model of individual decision making, an agent is faced with a choice between a number of different feasible *actions*. Each action can, in turn, have a number of different possible *outcomes*, where the actual outcome depends on which of a number of different possible *states of the world* obtains. (Formally, an action is a function from states of the world to outcomes, which maps each state to the outcome of taking the action in that state.)

Now suppose the agent has a *utility function* over the set of possible outcomes (which assigns to each outcome the agent's utility under that outcome) and a *credence function* over the set of possible states of the world (which assigns to each state its subjective probability, according to the agent's beliefs). Then the agent can in principle calculate his or her expected utility for each of the feasible actions. If the agent is rational, he or she will perform the action that maximizes expected utility.

[14] See Frankfurt (1969).
[15] See, e.g., Savage (1954).
[16] See, e.g., Dietrich and List (forthcoming).

This simple decision-theoretic model can be naturally embedded in the formal architecture we have suggested for thinking about free will. The agent's state at the time of the decision is given by his or her utility and credence functions (and possibly by other relevant properties). The agent's history at the time of the decision exhibits a branching point. Different branches correspond to the different feasible actions the agent could take. Each feasible action thus counts as *agentially possible*. Yet only those branches that correspond to actions that maximize the agent's expected utility count as endorsed. Formally, the *endorsement function* maps the agent's truncated history up to the given time to the set of those continuations (possibly singleton) that maximize the agent's expected utility according to the utility-credence-function-pair constituting the agent's state at the branching point.

In this way, decision theory accommodates *both* the possibility of acting otherwise *and* the existence of an endorsed action (or set of actions). While many actions may be feasible, only some maximize expected utility. An agent in standard decision theory can thus have free will in both a negative and a positive sense: he or she has the ability to act otherwise (provided there is more than one feasible action), and, if rational, acts with endorsement. This completes the simple, illustrative concretization of our framework.

5. Some complications that can undermine an agent's ability to act with endorsement

While our model shows that the ability to act otherwise is compatible with the ability to act with endorsement (though not generally with the ability to act otherwise with endorsement), there are also cases in which, even in the presence of alternative possibilities, the agent's ability to act with endorsement may be compromised. In such cases, the agent is free in what we have described as a negative sense, but not in its complementary, positive sense. We now discuss four cases of this kind.

5.1 No endorsed course of action at a particular branching point

While in a well-behaved case any truncated history $\mathbb{h}_t$ at a branching point will have at least one endorsed continuation, this need not be so in general. As mentioned above, the endorsement function might assign an *empty* set of endorsed continuations to $\mathbb{h}_t$. None of the continuations available at this branching point would then be endorsed by the agent: all possible actions would be unacceptable from his or her point of view.

We can refer to such cases as *agential dilemmas*, in analogy to moral dilemmas, in which every available course of action is morally wrong. Sophie's choice from the well-known novel by William Styron (1979) may be an extreme example of an agential dilemma. When Sophie has to choose which of her children should be saved and – by implication – which should be allowed to be killed, she cannot plausibly endorse any of the options, even though she must choose one in order to prevent the outcome that both of her children are put to death.

5.2 No fact of the matter as to which course of action is endorsed

A different kind of problem arises at branching points for which the endorsement function is *undefined*. Formally, this happens if the relevant truncated history $\mathbb{h}_t$ is not included in the domain of the endorsement function. While in an agential dilemma it is true of every continuation that it is

not endorsed, in the present case the question of whether a continuation is endorsed or not has no answer. There is no fact of the matter as to whether the agent endorses any given continuation of his or her history or not. Both the positive and the negative endorsement claims are indeterminate given the state of the agent at time *t*. Choice situations outside the domain of what can be rationally adjudicated – if such situations exist – might give rise to such indeterminacies.[17]

5.3 Conflicts between different modes of endorsement

Yet another possibility is that in some cases an agent finds him- or herself torn between different reasons for or against the various actions (or between different ways of weighing them), without being able to arrive at a balanced view. To represent cases of this kind, a more complex model may be required.

One modelling option would be to redefine an endorsement function as a mapping that assigns to each truncated history in its domain not a set of endorsed continuations but a family of such sets, with each set in the family being supported by one group of reasons (or by one way of weighing them). The different sets in the family might have some continuation(s) of the given truncated history in common, which would allow the agent to make a "safe" choice, without having to resolve the underlying conflict of reasons, or they might not. In the latter case, the agent faces a problem that is similar, to some extent, to an agential dilemma.

Another modelling option would be to keep the original definition of an endorsement function in place (i.e., to define it as a mapping from truncated histories to sets of endorsed continuations, rather than to families of such sets), but to allow that there might be more than one such function at work. The state of an agent who is torn between different reasons could then be described in terms of a "rivalry" between different endorsement functions, which pick out different and possibly disjoint sets of continuations of the agent's truncated history.

5.4 Intertemporal inconsistencies in endorsement

A well-behaved endorsement function may be expected to have the property of *intertemporal consistency*: other things being equal, if a particular continuation of an agent's history is endorsed at an earlier branching point, it will continue to be endorsed at later branching points, so that there is no need for the agent to deviate from previously endorsed continuations of his or her history:

> **Intertemporal consistency:** For every truncated history $\mathbb{h}_t$ in the domain of the endorsement function, if $\mathbb{h}'$ is an endorsed continuation of $\mathbb{h}_t$, then, for every later point in time t' for which $\mathbb{h}'_{t'}$ is also in the domain of the endorsement function, $\mathbb{h}'$ remains an endorsed continuation of $\mathbb{h}'_{t'}$.[18]

[17] In a more complex model, we might allow an endorsement function to map each truncated history in its domain to a three-fold partition of the set of its possible continuations: one segment of the partition would contain those continuations that are definitely endorsed; another segment would contain those that are definitely not endorsed; and the third segment would contain those for which it is indeterminate whether they are endorsed or not.

[18] Formally, for any truncated history $\mathbb{h}_t$ in the domain of e, if $\mathbb{h}' \in e(\mathbb{h}_t)$ then $\mathbb{h}' \in e(\mathbb{h}'_{t'})$ for any time t' after t with $\mathbb{h}'_{t'}$ in the domain of e (where e is the endorsement function).

Although intertemporal consistency is a nice property, there is no guarantee that the agent's endorsement function will always satisfy it. For this reason, there is room for another way in which the agent's ability to act with endorsement may be undermined. Suppose that the agent's set of endorsed continuations of h_t is non-empty, but every continuation h' in it sooner or later ceases to be endorsed: at some later time t' at which a branching point occurs in h', h' will not be among the endorsed continuations of $h'_{t'}$. If the agent can already predict this at time t, he or she may well experience the choice situation as problematic: choosing an endorsed action now may preclude the choice of an endorsed action later.[19]

To summarize, we have considered four kinds of situation in which the agent's ability to act with endorsement is undermined:

(i) agential dilemmas, in which the set of endorsed continuations of the agent's history at a particular branching point is empty;
(ii) indeterminate cases, in which the endorsement function is undefined at a given branching point, so that there is no fact of the matter as to which action the agent endorses;
(iii) unresolved conflicts between different reasons, where the agent is torn between different sets of endorsed continuations of his or her history or between different endorsement functions; and, finally,
(iv) intertemporal inconsistencies, where the agent can predict that none of the endorsed continuations of his or her history will continue to be endorsed in the future.

6. Endorsement as an intensional notion

So far, we have assumed that what is endorsed at any branching point in a given history is one or several *concrete histories*: (some of the) continuations of the truncated history at this point. However, this way of conceptualizing endorsement might be accused of being too "extensional". What we do endorse, an objector might say, is not concrete histories but propositions that describe these histories, and the actions performed in them, in more or less detail. Such descriptions cannot be fully specific, given our limited abilities to grasp the future. From this "intensional" perspective, an endorsement function should map truncated histories not to sets of endorsed *histories*, but to endorsed *propositions*.[20]

The first thing to note is that even our current extensional definition of an endorsement function admits a propositional interpretation. This is because truncated histories are mapped to *sets* of histories, which, in turn, can be interpreted as propositions. (Recall that any subset of Ω, the underlying set of all possible agential histories, can be interpreted as a proposition, which is true in

[19] Another issue should be mentioned here. Even when the endorsement function is intertemporally consistent, it is still important to define it also for branching points that lie "off any equilibrium path", so to speak, i.e., even for branching points that are reachable only if the agent at some earlier branching point chooses an action he or she does *not* endorse. Especially in game-theoretic contexts, such "off equilibrium" choices are important to consider, e.g., when one determines the backward-induction solution for an extensive-form game with perfect information by making predictions about what the players would choose at various future branching points, *if* they were to reach them.

[20] Another alternative, already mentioned in relation to the example of Luther, would be to individuate histories in a very coarse-grained way, i.e., to look at them as essentially abstract rather than concrete objects. Such histories could be objects of endorsement, even if one interprets endorsement as an intensional notion.

all histories contained in it, and false in all others.) So, one might say, under the existing definition any truncated history is already mapped to an endorsed proposition.

Nonetheless, there are other, more general ways of constructing an *intensional endorsement function*. The general strategy is to define it as a function

- whose domain, as before, is a set of truncated histories,
- and whose co-domain is *either* (i) a set of propositions, *or* (ii) a set of sets of propositions,

where

> any truncated history in the relevant domain is mapped to (i) the proposition or (ii) the set of propositions that the agent "endorses" – or, in alternative terminology, "rationally supports".

Note that the two alternative ways of defining the output of the function – either as a single proposition or as a set of propositions – are formally distinct. Even when we can represent the set of propositions in the second case as a single conjunctive proposition (which is possible so long as the set contains only finitely many propositions), any such translation involves some informational loss. So let us look at different ways in which we might specify and interpret an intensional endorsement function, and ask how these relate to extensional endorsement functions as defined earlier.

Consider first case (i), where the output of the intensional endorsement function for each truncated history is a single proposition. One might think that we have already covered this case. As noted, an extensional endorsement function maps any truncated history to a set of endorsed histories, which in turn can be seen as a proposition: the proposition that is true in precisely those histories that are endorsed. So there appears to be a one-to-one correspondence between extensional and intensional endorsement functions, assuming the latter produce single propositions, not sets of propositions, as output. The appearance of this one-to-one correspondence, however, is misleading. Since an extensional endorsement function maps any truncated history $\mathbb{h}_t$ to a set of continuations of $\mathbb{h}_t$, the proposition it thereby picks out is true only in continuations of $\mathbb{h}_t$, not in any other histories. This is a significant restriction. More plausibly, the agent's object of endorsement may be a proposition in general, which picks out *some* subset of $\mathbb{\Omega}$; this may intersect with the set of continuations of $\mathbb{h}_t$ but need not consist solely of such continuations.

We can then retrieve an extensional endorsement function from an intensional one as follows. For each truncated history $\mathbb{h}_t$ (the input of both functions), the set of endorsed histories (the output of the extensional endorsement function) is the intersection of the endorsed proposition (the output of the intensional endorsement function) and the set of possible continuations of $\mathbb{h}_t$. Since different propositions can have the same intersection with the set of possible continuations of $\mathbb{h}_t$, intensional endorsement functions stand in a many-to-one relation to extensional ones: different intensional endorsement functions can give rise to the same extensional endorsement function.

Note that this opens up two ways in which the agent's set of endorsed histories at a given time may be empty, so that an agential dilemma arises. It may be empty because the agent's endorsed proposition is an impossible one, represented by the empty set and thus false in all histories. Or it may be empty because the endorsed proposition, while represented by a non-empty subset of $\mathbb{\Omega}$,

happens not to overlap with the set of continuations of the agent's truncated history. In this case, the endorsed proposition is not impossible by itself, but impossible in the given context.

Case (ii), where the output of the intensional endorsement function is a set of propositions, rather than a single proposition, can be analyzed similarly. Here any truncated history $\mathbb{h}_t$ is mapped to the set of propositions that the agent endorses at time *t*. Intuitively, these are the propositions the agent seeks to make true through any endorsed action at time *t*. It should be evident that, as in case (i), any intensional endorsement function in case (ii) also determines an extensional endorsement function, where the determination relation is once again many-to-one. Now the set of endorsed histories for any truncated history $\mathbb{h}_t$ is the set of those continuations of $\mathbb{h}_t$ in which *all* of the endorsed propositions are true. Equivalently, it is the intersection of all the endorsed propositions and the set of possible continuations of $\mathbb{h}_t$.

There are now four ways in which the set of endorsed histories at a given time may be empty. It may be empty because some proposition among the endorsed propositions for a truncated history $\mathbb{h}_t$ may be impossible (i.e., empty) or impossible in a given context (i.e., non-overlapping with the set of continuations of $\mathbb{h}_t$). Or it may be empty because the different endorsed propositions, while separately possible to realize, are mutually incompatible. Either they cannot be realized together at all, or they cannot be realized together in a given context. In the latter case, there are histories that instantiate all the endorsed propositions, but none of them is a continuation of $\mathbb{h}_t$.

6. Concluding remarks

We have developed a simple model to show that two intuitions about free will can be reconciled: the intuition that free will requires the ability to do otherwise, and the intuition that this ability alone is insufficient for free will, since free will requires more than making un-determined fluke choices. The central observation has been that there is an important distinction between *possible* and *endorsed* actions. Our model has offered conceptual resources for capturing both notions and for analyzing their relationship to one another.

In line with other contributions to the literature, we have argued that there are at least two senses in which an action can be free: a "negative" sense, which ties freedom to the ability to do otherwise, and a "positive" sense, which ties freedom to endorsement. Our analysis shows that one can *consistently* hold that free will requires *both* the ability to act otherwise *and* the ability to act with endorsement, but that one cannot generally (but at most in special cases) hold that free will requires the ability to act otherwise with endorsement. Recognizing this last point is crucial for resolving the tension between the two rival intuitions we have started with.

Although we have defended a formal model for thinking about free will, we have tried to remain as neutral as possible in filling in the interpretation of the various elements of that model. In analyzing an agent's possibility of doing otherwise, we have not taken a view on how exactly *agential possibility* is related to *physical* or *metaphysical possibility*. Both libertarians, who typically employ a physical or metaphysical notion of possibility, and those compatibilists who regard agential possibility as distinct from physical or metaphysical possibility should be able to use our model. And in defining an endorsement function, we have tried to stay neutral between different philosophical views on what endorsement might consist in. Our goal has been to discuss the relationship between two initial intuitions about free will from a general, purely structural

perspective. This should allow our model to serve as a framework in which different substantive theories of freedom can be located and compared.

References

Belnap, N. D., M. Perloff, and M. Xu (2002) *Facing the Future: Agents and Choices in our Indeterminist World*, Oxford (Oxford University Press)

Belnap, N. D. (2005) "Branching Histories Approach to Indeterminism and Free Will", in Bryson Brown and Francois Lepage (eds.), *Truth and Probability: Essays in Honour of Hugues Leblanc*, London (College Publications), pp. 197-211

Berlin, I. (1958) *Two Concepts of Liberty*, inaugural lecture, Oxford (Clarendon Press)

Dennett, D. (1984) *Elbow Room*, Cambridge (Cambridge University Press)

Dennett, D. (2003) *Freedom Evolves*, London (Penguin)

Dietrich, F., and C. List (forthcoming) "A reason-based theory of rational choice", *Nous*

Fischer, J. M., R. Kane, D. Pereboom, and M. Vargas (2007) *Four Views on Free Will*, Oxford (Blackwell)

Frankfurt, H. (1960) "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66: 829-839

Kane, R. (ed. with introduction) (2002) *The Oxford Handbook of Free Will*, Oxford (Oxford University Press)

List, C. (2011) "Free will, determinism, and the possibility of doing otherwise", working paper, London School of Economics (http://personal.lse.ac.uk/LIST/FreeWill.pdf)

Maier, J. (2011) "The Agentive Modalities", working paper, Australian National University

Pereboom, D. (2005) "Defending Hard Incompatibilism", *Midwest Studies in Philosophy* 29: 228-247

Savage, L. (1954) *The Foundations of Statistics*, New York (Wiley)

Segerberg, K. (1984) "Towards an Exact Philosophy of Action", *Topoi* 3: 75-83