

FREE WILL AND RATIONALITY

António Zilhão
University of Lisbon

1. Incompatibilism and Libertarianism

Incompatibilism is the doctrine according to which the truth of determinism entails that there is no free will. One of its versions - libertarianism - claims that there is free will. Libertarians thus conclude, by *Modus Tollens*, that determinism must be false.

Typically, the incompatibilist premise of the libertarian argument is argued for by means of philosophical reasoning. Usually, this reasoning proceeds along the steps of the so-called *Consequence Argument*. These are the following. First, the statement that the truth of determinism entails that our actions are a consequence of both events that occurred long before we were born and the laws of nature. Second, the statement that both what happened before we were born and the laws of nature are entirely beyond our control. Third, the statement that whatever is a consequence of these past occurrences and the laws of nature must also happen entirely beyond our control. Fourth, the statement that whatever else our having free will might mean it certainly must mean that we do have control over our actions. Finally, the conclusion that, if determinism is true, then we have no free will.

Regardless of the question of how convincing this chain deduction really is (Dennett, for instance, produced an interesting evolutionary counter-argument against it), it is meant to justify only the major premise of the libertarian argument. Obviously, determinist hardliners accept it too. Contrary to libertarians though, they affirm that determinism is indeed true. This they typically justify by appealing to what they take to be the current scientific world view, deemed to be generally accepted everywhere outside of quantum mechanics. Therefore – they conclude by *Modus Ponens* – we have no free will.

What confers its distinctive character to the libertarian position within incompatibilism is thus the minor premise of the *Modus Tollens* argument by means of which it is arrived at. As mentioned above, this premise states that there is free will (because we have it). But how do libertarians know this to be the case?

The free will libertarians claim we do have is of a peculiar sort. This makes the question above all the more relevant. As a matter of fact,

libertarian free will is conceived of as satisfying a particularly strong reading of two of the defining principles of this concept. These are the principle of alternative possibilities (PAP) and the principle of ultimate responsibility (PUR). The former states that, in order to enjoy of free will, the agent must have beforehand a *genuine* possibility of choosing from among different available courses of action; and the latter states that, in order to enjoy of free will, the agent must bear *ultimate* responsibility for the action he ends up performing. According to the libertarian reading of these italicized terms, in order to satisfy any of these principles the free agent must be a sort of Aristotelian unmoved prime mover with respect to his actions. Now, I, and presumably lots of other people, we certainly need to be persuaded that something of this sort might be true of us.

2. Justifying the Minor Premise

Surprisingly, there is little agreement on how to answer the question above within libertarian literature itself. Some libertarian philosophers (such as, e.g., K. Lehrer) claim nothing else is needed in order to support the minor premise of the libertarian argument besides our having a very powerful intuition that it is indeed true. But is it appropriate to claim that it is true that we are free-willed because we (or some of us) have the intuition that we are? I take this view to be highly implausible. Others seem to think the same. Within the libertarian camp, a major philosopher holding an anti-intuitionist view is van Inwagen. Indeed, he explicitly claims that a justification of the minor premise of the libertarian argument simply in terms of the appeal to the intuition that it is true cannot be accepted. However, he purports both to sustain this claim and to defend his own libertarian alternative to it by means of a combination of philosophical arguments I find rather awkward. Let me try to show you why.

First, let us consider what we might call his *anti-intuitionist* argument. The first step of this argument consists of a thought-experiment. This is the following. Imagine that, when any human being is born, a device is remotely implanted in his brain by some clever race of extraterrestrians; and imagine further that this device is such that, whenever any human must make a decision, it causes him to have the set of beliefs, desires and intentions that will lead him to decide one way or another in agreement with the prescriptions of a decision-making program created by the extraterrestrians; moreover, this device is supposed to be such that it is not detectable by any of the brainscan techniques we currently possess. Under such imaginary circumstances, van Inwagen claims, humans would obviously not enjoy of free will; as a matter of fact, they would be nothing but remote controlled puppets.

Van Inwagen wants us then to consider the matters this thought-experiment describes from the standpoint of our experience. Once we do that, we are presented with the following argument. First, we should

recognize that, viewing things from this perspective, the possible world envisaged in this thought-experiment is indistinguishable from the world we think we live in. Secondly, having recognized this, we are led to agree that, in that possible world, we would certainly have the same intuitions we do have in the supposedly actual world; one of these is, of course, the intuition that we are free-willed; thus, we would have it there too. Thirdly, irrespective of how powerful this intuition might be in this possible world, we have to concede that it would there be false. The fourth step of the argument is then the claim that it follows from what was already acknowledged in the first step that we cannot prove that the world we live in is not in reality the possible world imagined in this thought-experiment. Finally, the conclusion of the argument states that we cannot infer the truth of the proposition that we are free-willed from the strength of our intuition that we are indeed free-willed.

Let us now consider his second, positive, argument. As mentioned above, being a libertarian, van Inwagen endorses the minor premise of the libertarian argument. Thus, he needs a non-intuitionist justification for this endorsement. Indeed, he claims not only that libertarians need an argument for this but also that he has found a good one. This argument consists of another *Modus Tollens* inference. Its major premise is the conditional proposition that if we do not have free will then we are not morally responsible; the minor premise is the proposition that we are, in fact, morally responsible; and the conclusion is, obviously, the proposition that it is not the case that we do not have free will. So, we have it.

Now, the idea that having free will, in some sense of this term, is a necessary condition for our having moral responsibility, in some sense of this other term, seems pretty straightforward. The problem for libertarian philosophers is, of course, proving that the concept of free will which is necessary for ascriptions of moral responsibility to make sense needs to be defined in terms of the satisfaction of the particularly strong reading of the PAP and of the PUR they endorse. But let us put this problem aside for the moment and concentrate on the issue that is here of crucial importance, namely, the following. How does van Inwagen justify his claim to know that we do enjoy of moral responsibility? Well, his answer to this question is basically the following: our life in common is such that we “cannot but view our belief in moral responsibility as a justified belief, a belief that is simply not open to reasonable doubt.”

Within the general economy of van Inwagen’s inferential construction, this is a startling answer to the question above, to say the least! As a matter of fact, I suppose no one fails to recognize in van Inwagen’s previously described anti-intuitionist argument a rather familiar form. It is the form characterizing one of the most famous philosophical arguments ever, namely, Descartes’ argument of the *malin génie*. In reality, the latter is, in turn, just a rewording of an even older argument, namely, the skeptic argument against the admissibility of inferring the reality of the outside world from the data of our experience. Now, this is

not the place to judge the merits or demerits of epistemic skepticism. But if an argument like van Inwagen's anti-intuitionist argument, modelled on the classical argument of epistemic skepticism, is supposed to be a good argument against the admissibility of the claim that we enjoy of free will because we have the intuition that we do, then how can the claim that we have free will because we have moral responsibility, and that we have moral responsibility because our belief that we do is beyond any reasonable doubt, be any *better*?

Let me press this point a bit further. It seems to me obvious that, whatever else belongs to the concept of moral responsibility, it only makes sense to claim that such a concept applies truly to us if we conceive of ourselves as being part of a world populated by communities of agents living in close interaction with each other. However, we can only claim to know we live in such a world if we have somehow managed to get over the skeptical arguments against the existence of the outside world and other minds. Thus, if the reason why we cannot rely on the truthfulness of our intuition that we have free will is that we don't know how to overcome skeptic arguments against the external existence of the objects of our cognition, the very same reason should prevent us from the outset from accepting the inference of the reality of moral responsibility from our belief in it as the foundation for the claim that we have free will.

Van Inwagen is thus being incoherent here. Indeed, I believe that, if he wants us to take seriously his argument against free will-intuitionism, then he cannot possibly want us to take seriously his foundational argument appealing to moral responsibility. Alternatively, if he wants us to take seriously his foundational argument appealing to moral responsibility, then he cannot possibly want us to take seriously his argument against free will-intuitionism.

Now, I do not think that skeptic arguments of the same form of those concerning the existence of the outside world and other minds are particularly relevant in a debate about the reality or unreality of free will. Or, at least, I do not think they are any more relevant here than they in general are for the assessment of any contention whatsoever concerning the scope and validity of our knowledge claims. So, let me ignore them here.

This said, don't misunderstand me. In spite of having disregarded epistemic skepticism, I do not consider that our alleged intuition concerning our being endowed with free will is at all of the same epistemic nature as are our intuitions that we are not brains in a vat, or that our own bodies exist in an independent material world, or that the two hands with which this paper is being typed are indeed my hands. Specially, if the free will which happens to be the target of this intuition is to be defined in terms of the above-mentioned strong reading of the two also above-mentioned principles, as is the case in libertarian thinking. Thus, I do not disagree with van Inwagen's claim that we do need an

argument in favour of the truth of the minor premise of the libertarian argument, if we are to give it some credit. I simply disagree with him on the appropriateness of the argument he puts forth in order to claim that we do need such an argument.

So let me choose the second prong of the alternative mentioned above, and let me take seriously van Inwagen's argument according to which it is true that we have free will because our having it is a necessary condition for our having moral responsibility and our belief that we have the latter is beyond any reasonable doubt; as a consequence, let me disregard van Inwagen's 'anti-intuitionist' argument.

3. 'Legislating for the universe on the basis of the present desires of men'

Now, I think there is a strong counter-argument to the use of our having moral responsibility as a minor premise in an argument such as van Inwagen's. As a matter of fact, van Inwagen himself considers this counter-argument, even if only to dismiss it. It is the following. If we were to accept van Inwagen's argument as a good argument, then we would be inferring substantial conclusions about how the physical world is supposed to be, namely, that determinism is false about nature in general, from premises having to do with moral notions, namely, the notion that we enjoy of moral responsibility. But the way the physical world is is surely totally independent from any details concerning the existence or inexistence of moral constraints on our actions. After all, the issue whether determinism is true or not would make sense even if the human race had never populated the earth and no moral notions ever came to exist.

This very same point was forcefully made by Bertrand Russell one century ago. He wrote, having Spencerian evolutionists in mind, that "to regard ethical notions as a key to the understanding of the world is essentially pre-Copernican". And to this he added: "human ethical notions (...) are essentially anthropocentric, and involve, when used in metaphysics, an attempt, however veiled, to legislate for the universe on the basis of the present desires of men." (Russell 1914, p. 83) Although written within a different context, these words may be taken to conclude appropriately what is essentially an argument carrying a considerable amount of inductive force.

Awkwardly, van Inwagen claims that an incompatibilist should answer such a counter-argument by saying that it carries no more force against his own argument than does a 'Universal deceiver' argument for skepticism. Thus, an incompatibilist "ought not be more troubled by this charge" than he is troubled by the latter sort of argument. Van Inwagen's is thus a strange sort of dialectic. In the course of one and the same general argument, he manages to produce, first, a sub-argument of the

'Universal deceiver' kind in order *to prove* that one cannot simply rest the claim that we do have free will on the intuition that we do have it, and, secondly, a swift dismissal of an inductively relevant argument against his own proposal of supporting the free will claim on the strength of our intuition that we are morally responsible by claiming that such a counter-argument has no more force against his proposal than has a skeptic argument of the 'Universal deceiver' kind. This is rather baffling; I simply do not know how to make sense of this peculiar conjunction of arguments. So let me move away from van Inwagen's version of libertarian anti-intuitionism and let me search for another non intuitionist way of justifying the minor premise of the libertarian argument.

4. Free Will as a necessary condition for Rationality.

In a recent book, suggestively titled *Rationality + Consciousness = Free Will*, David Hodgson claims that the alleged truth of the minor premise of the libertarian argument admits being derived from our being rational conscious beings. Obviously, neither the concept of rationality nor the concepts of consciousness or free will belong to the sort of entities of which it could make sense to conceive of them as being either the arguments or the values of an arithmetical function like addition. In reality, what Hodgson's view may be taken to illustrate is a standpoint according to which the having of free will is, among other things, a *necessary* condition for truthful ascriptions of rationality and consciousness.

Now, let us suppose this contention is accepted. Under such circumstances, the contrapositive of the statement expressing it can be taken to perform the role of the major premise in a *Modus Tollens* inference, the minor premise of which would be the proposition that we are, as a matter of fact, conscious rational beings; the conclusion of this inference would obviously be the proposition that it is not the case that we do not have free will. This would be formally similar to van Inwagen's above described argument, but the concepts of rationality and consciousness would replace in this alternative argument the role the concept of moral responsibility played in the original one. Let me then take a look at this alternative both to libertarian intuitionism and to van Inwagen's anti-intuitionist approach.

I do not doubt that we are conscious. I take it this goes without saying. Whether or not having consciousness might be necessary for having free will is, however, another matter. But it is the other term of the left hand side of Hodgson's equation I want to focus on. Namely, I want to discuss whether it is really the case that rationality might play the role Hodgson ascribes to it, i.e., that of being a necessary part of a sufficient condition for free will. Indeed, most participants in the free will debate seem to assume that no other mortal creatures besides us, the rational animals,

are free-willed; this points to the existence of a presumably relevant link connecting rationality and free will. A possible justification for the existence of this link could be the following: free will emerges from a specific kind of cognitive complexity, namely, the kind of cognitive complexity that also subserves the emergence of rationality in us.

Being a necessary part of a sufficient condition for free will, truthful ascriptions of rationality could then play the demarcation role typically played by an adequacy condition. They would then separate the purposeful behaviour of cognitively sophisticated agents of whom free will could be meaningfully ascribed from the purposeful behaviour of less sophisticated creatures of whom free will could not be meaningfully ascribed.

But what it is that one ascribes, when one is truthfully ascribing rationality to a creature? In his above-mentioned book Hodgson never defines this term. He concentrates his attention rather in what he sometimes calls “instinctive informal rationality”, a concept the content of which is actually far from clear. So, let me put aside Hodgson’s own *sui generis* approach to this topic, and let me try to figure out how a more traditionally defined concept of rationality would fare in the above-mentioned role.

5. Aristotle, the Fable of the Bees and Folk-Psychological Law

According to philosophical tradition, rationality is what is possessed by a rational agent; a rational agent is, in turn, a creature α that behaves according to the following scheme (I+II) – the so-called ‘practical syllogism’, first introduced by Aristotle:

(I):

α has a desire δ the content of which is γ ;
 α has a belief β the content of which is that doing θ is the best thing to do to get γ ;

$\therefore \alpha$ believes that doing θ is the best thing for him to do.

(II)

α believes that doing θ is the best thing for him to do.
 α believes that no obstacle prevents him from doing θ .

$\therefore \alpha$ does θ .

Two millennia after Aristotle, this scheme kept providing the model for the description of the behaviour of rational agents (typically, persons).

This much may be attested by reading the following words Bernard de Mandeville wrote in his most famous essay *The Fable of the Bees: or Private Vices, Publick Benefits*:

“no Person can commit or set about an Action, which at that then present time seems not to be the best to him.” (vol. II, p. 196).

and

“When two Things are left to a Person’s Choice, it is a Demonstration, that he thinks That most eligible which he chuses, how contradictory, impertinent or pernicious soever his Reason for chusing it may be: Without this there could be no voluntary suicide; and it would be Injustice to punish Men for their crimes.” ” (vol. II, p. 197).

Later on, in the 1970s, the philosopher Paul Churchland proposed the following multiply quantified conditional sentence as a contemporary update of traditional Aristotelian rationality:

“(∀X)(∀Φ)(∀A){[1.(X wants Φ)& 2.(X believes that A-ing is a way for him to bring about Φ under the existing circumstances) & 3.(there is no action believed by X to be a way for him to bring about Φ, under the circumstances, which X judges to be as preferable to him as, or more preferable to him than, A-ing) & 4. (X has no other want or wants which, under the circumstances, overrides his want to Φ) & 5.(X knows how to A) & 6. (X is able to A)] → (X A-s)}”.

More importantly, Churchland dubbed this sentence to be the nomic kernel of what, in philosophical jargon, was to become known as ‘Folk-Psychology’, that is, the proto-scientific theory we all use when trying to understand, *explain* and predict the behaviour of our fellow humans (the rational animals).

6. Newell’s Cognitive Systems

However, at least since the mid-1950’s, the very same scheme of the practical syllogism came also to be adopted to serve as a definition of a different and unabashedly deterministic notion – the notion of a cognitive system. Allen Newell, a distinguished researcher in AI, characterized this notion in more or less the following terms. A cognitive system is a physical system the behaviour of which obeys the following general law: if the system has a particular goal *O* and if it harbours the knowledge items *C*₁,...,*C*_{*n*}, according to which performing action *A* is conducive to the bringing about of goal *O*, then the system does action *A*.

When introducing the notion of a cognitive system, Newell was talking about all kinds of purposeful physical systems including, among others, animals, even very primitive ones, and robots and other complex

machines. No assumption of free will is made in his definition above, and none should be.

Obviously, accepting that Newell's notion is well motivated is tantamount to accept that rationality, thus understood, is perfectly independent from free will. Thus, after Newell and AI, if rationality were to remain being understood in the Aristotelian way, its use as an adequacy condition for ascriptions of free will would be bogus.

7. What makes best best?

But the definition of what is a rational agent did change in interesting respects.

As a matter of fact, more or less at the same time AI was giving its first steps, the Aristotelian definition of rationality was taken by many people in the field to be too coarse to adequately define the behaviour of a rational agent. In particular, people felt that the notion of an agent taking a particular course of action to be the best, given his desires and his beliefs about the world, was still in need of clarification.

Underlying this need was the realization that, given the same set of desires and beliefs about the world, different courses of action were always available to the agent; however, assuming that 'best' did not mean simply 'what happened to be chosen', there had to be some objectively identifiable criterion in terms of the satisfaction of which one of the available options admitted being selected as subjectively better than the others.

The criterion of optimality that ended up being proposed was the so-called principle of 'maximization of expected utility'. Adopting this principle as a criterion of rationality meant that the choice behaviour of a rational agent could be appropriately modelled by the following set of procedures:

1. the agent's beliefs about the obtaining of each of the relevant possible states of the world were to be measured by a probability function;
2. the numerical values thus obtained were to be multiplied by the numerical values measuring the comparative utility or desirability of each of the possible outcomes of the action under consideration;
3. the products so obtained were to be added;
4. the sum thus obtained would exhibit a value – called the 'expected utility' of undertaking a given course of action – that would allow the agent to numerically compare it with the values of the other courses of action available to him;
5. a decision to act in a certain way were to be reached by selecting the course of action enjoying of the highest expected utility.

The adoption of this criterion to be the mark of rationality is not arbitrary. It is based upon the analysis of the principles underlying what admits being called 'rational betting behaviour'. This is the behaviour someone involved in a betting game has to follow in order not to suffer a guaranteed loss at the game.

Thus, the claim that rational agency is action in order to maximize expected utility is a claim that can be mathematically proven once one accepts the principle that the general behaviour of a rational agent is to be modelled upon the behaviour of a rational better in a betting game. In fact, this claim is a *theorem* that follows from the axioms in terms of which the latter behaviour admits being formalized.

8. Pigeons do reasonably well in conforming to the axioms of rational-choice theory

Having gone through this clarification of what it means to truthfully ascribe rationality to an agent we are now in a position to ask the following question: is the possibility of modeling our purposeful behaviour as the behaviour of rational betting agents the watermark that identifies the kind of peculiar cognitive complexity we are endowed with?

In particular, is it what identifies the kind of cognitive complexity that might also be responsible for our being somehow endowed with free will? Let us examine this.

The title of this section refers to a paper by C. J. Kagel published about twenty-five years ago. It expresses fairly clearly the general idea of an interesting new area of research. Since then, a whole body of literature has grown devoted to the theme of successfully showing not only that animal behaviour in general admits being interpreted as satisfying the axioms of rationality, but that it fares *better* than human behaviour in that respect (other suggestive titles in this literature are "Are humans less rational than lower animals?" or "Rational animals?" or "Rationality in risk-sensitive foraging choices by starlings").

Simultaneously, an enormous amount of literature has been published claiming that human behaviour does poorly in the business of being rational, i.e., that humans often and unmistakably violate the axioms of rationality (by, *inter alia*, Tversky, Kahneman, Lichtenstein, Slovic, etc.)

There is something startling about this. How can the literature above be made to cohere with the work undertaken by rationality theorists, and the theoretical expectations contained in it?

Let's begin answering this question by taking a look at the axioms of rationality theory.

9. Axiom of Independence

One of the main axioms of rationality theory is the axiom of independence. It establishes the following: if x is chosen from the choice set x and y , then y cannot be chosen when the choice set is widened to x, y , and z .

To try to figure out whether or not independence should be an axiom of rationality imagine what would happen if independence were violated. The following story is a good example of this.

Imagine that you were invited for dinner at your friend's home. As a display of courtesy towards you, your friend invites you to choose the wine. 'I've a bottle of red (x) and a bottle of white (y); which one do you prefer?' he asks you. You say you prefer the red (x). He then says: 'Oh, I forgot. I also have a bottle of rosé (z).' You then reply: 'Oh, OK; in that case I prefer the white (y)'.

Your behaviour *does* seem weird. Your friend will probably think you're teasing him and won't take you seriously. Thus, taking this example as a standard of what would have to happen if people would go about violating the axiom of independence does seem to indicate that, under those circumstances, human life would become unrecognizably alien.

Thus, assuming independence seems to make sense. But does it?

Consider now another story. Keep imagining that you were invited for dinner at your friend's place. Imagine that, after the main course is over, your friend realizes that he forgot to buy desert. He has a single portion of apple pie in the fridge though; he knows you like apple pie and asks you if you want to have it. You know that he also likes apple pie. So, you decline, thus choosing nothing (x) over apple pie (y).

A few moments later, another friend joins you for coffee; he brings with him a set of chocolates from the deli around the corner. Offered to have some chocolate (z), you decline the offer and choose the apple pie (y) instead.

Seemingly, you did here just the same thing you did in the previous example – you violated the axiom of independence: you chose x from the choice set x and y , and then you chose y , when the choice set was widened to x, y , and z .

However, your behaviour does not appear weird anymore; not only will your friend not think you are teasing him but he will be moved by your attitude and take it as a token of your friendship.

10. Axiom of Cancellation and (un)sure things

Let's now consider another axiom: Cancellation or Substitutability. Basically, it states the following: if an option A is at least as preferred as an option B, then, if options C and D result from, respectively, options A and B by a common change in their outcomes, then option C is at least as preferred as option D. (i.e., states with the same outcomes should cancel out).

Again, that states with the same outcomes should cancel out in a choice process seems to make obvious sense. Why should they make a difference at all?

But is this really true? Let's consider a particular decision problem, devised precisely to test this axiom. This is the so-called 'Allais Problem'. It is the following:

1. Make a choice between the following two options: Option A: a gamble in which you win \$1,000,000 no matter what. Option B: a gamble in which you have a 0.89 probability of winning \$1,000,000, a 0.10 probability of winning \$5,000,000 and a 0.01 probability of winning nothing.
2. Make a choice between the following two options: Option C: a gamble in which you have a 0.11 probability of winning \$1,000,000 and a 0.89 probability of winning nothing. Option D: a gamble in which you have a 0.10 probability of winning \$5,000,000 and a 0.90 probability of winning nothing.

How did you choose?

Like most people, you probably chose option A over option B and option D over option C. But options C and D result from options A and B by a common change in their outcomes. That is, if you did choose that way, your choice behaviour admits being represented thus:

Problem 1: $0.11U(1,000,000) > 0.10U(5,000,000)$ and
Problem 2: $0.10U(5,000,000) > 0.11U(1,000,000)$.

Choosing this way is, however, choosing in a contradictory way.

According to the axiom of cancellation, in order to be rational, if you preferred option A over option B in the first choice you had to have preferred option C over option D in the second choice.

11. Going beyond immediate or 'objective' utilities

In both cases, it is possible to reintroduce consistency by reinterpreting differently the choice problems faced by the agents. In order to do that,

one could (should) argue that the situations devised above need to be contextualised beyond the immediate or 'objective' utilities presented.

In the case of independence, you could (should) say that, in your first choice, you were not simply preferring nothing to apple pie; you were preferring nothing (and being polite) to grabbing the only piece of dessert your friend had left in the fridge (and being unpolite and let everybody else hate you).

In the case of cancellation, you could (should) say that by being simultaneously risk-averse when a gain is the dominant outcome and risk-prone when a loss is the dominant outcome, you were actually maximizing avoidance of regret. Given that regret is a fact of human psychological life with unpleasant experiential consequences, it would make sense to go beyond the immediate utilities presented and incorporate the expectation of regret into the definition of the utility of the outcomes.

Is there a problem with the following of this strategy in order to reintroduce consistency?

Well, there is, actually. As Stanovich emphasized, the axioms require that you should abstract away from most aspects of the contextual environment of the choice-problems you face; but ordinary human life cannot avoid social contextualizing.

This tension, however, is absent when nonhuman animal behaviour is modelled in terms of the axioms of rational choice. Typically, nonhuman animals are responsive to immediate 'objective' utilities alone, independently of context.

In consequence, I would like to echo here the following suggestion put forward by Stanovich The empirical results reported in the literature mentioned above can be made to cohere with the prescriptions of rationality theory in the following way: beyond a fairly elementary level of complexity, the more complex the cognitive architecture of an organism is, the more difficult it is for it to be rational.

If this is indeed the case, then rationality can certainly not play the role of an adequacy condition for free will if we are to have it and non human animals are not to have it. On the contrary, it seems that the peculiar sort of cognitive complexity humans display puts them at a *disadvantage* when trying to be rational.

12. A two-way power indeed

In short, neither knowledge from moral responsibility nor rationality will successfully replace intuitionism about free will as a means of justifying

the minor premise of the libertarian version of the incompatibilist argument. Of course, this does not mean that other alternatives to it cannot be imagined. But it is certainly not easy to foresee what could possibly be the evidence supporting the claim that any such hypothetical alternative could play this justificatory role.

Let me now present, in closing, what I take to be a viable hypothesis for making sense of the content of libertarian intuitions about our being endowed of free will. Typically, libertarian intuitionists phrase their description of this content by means of the expression that we enjoy of a 'two-way power' with respect to the things that we deliberate upon. I do agree many of us have an intuition more or less corresponding to this description. However, and in view of the previous sections, I think it makes more sense to interpret the intuition that we enjoy of such a power as being an intuition generated by our awareness of the essential *unstableness* of our choice experiences. The latter is in turn the result of the difficulties our cognitive make-up lays down to our attempts at being rational under the complexity of the information that we seek to bring to bear on the decision processes we get involved into. Thus, rather than being generated by our experience of, in fact, exercising a mysterious god-like metaphysical power, such an intuition would be, according to my proposal, a reflection of our all too real and down to earth cognitive limitations.

Indeed, the fact that we are social contextualisers (and also second-order evaluators of our own first-order mental life, as Frankfurt has rightly stressed) makes the rational framing of *any* decision of ours a highly precarious business. This is why, I think, we systematically *feel* that we could have done otherwise under the very same circumstances (which is, in fact, the phenomenological content associated both with the expression stating that 'we have a two-way power with respect to the things that we deliberate upon' and with the intuition that a strong reading of the principle of alternative possibilities is satisfied in our choices).

Of course, if I am right in this respect, the libertarian that backs on his having such an intuition the minor premise of the argument defining his version of incompatibilism is incurring in a semantic fallacy. As a matter of fact, he is implicitly interpreting the minor premise so as to be plausibly empirically true, and seems not to realize that, under the circumstances under which it is plausibly empirically true, the reference of the term 'free will' in the minor premise is not the same as the reference of the same term in the major premise.

LITERATURE:

Allais, M. 1953, "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine".
Econometrica, 21, 503-546.

- Aristotle 1941, *Nicomachean Ethics*. In *The Basic Works of Aristotle* (ed. by McKeon. New York: Random House.
- Churchland, P. 1970, "The Logical Character of Action-Explanations". *The Philosophical Review*, 79, 214-236.
- Dennett, D. 2003, *Freedom Evolves*. London: Alan Lane.
- Frankfurt, H. 1971, "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy*, 68, 5-20.
- Frankfurt, H. 2009, *The importance of what we care about – Philosophical essays*. Cambridge: Cambridge University Press.
- Hodgson, D. 2012, *Rationality + Consciousness = Free Will*. Oxford: Oxford University Press.
- Hurley, S. & Nudds, M. (eds.) 2006, *Rational Animals?* Oxford: Oxford University Press.
- Jensen, K., Call, J. & Tomasello, M. 2007, "Chimpanzees are rational maximizers in an ultimatum game. *Science*, 318, 107-109.
- Kagel, C. J. 1987, "Economics according to the rats (and pigeons too): What we have learned and what we hope to learn". In Roth, A. (ed.), *Laboratory experimentation in economics: Six points of view*. Cambridge: Cambridge University Press, 587-703.
- Lehrer, K. 1960, "Can we know that we have free will by introspection?". *The Journal of Philosophy*, 57, 145-157.
- Lichtenstein, S. & Slovic, P. (eds.) 2006, *The construction of preference*. Cambridge: Cambridge University Press.
- Kahneman, D. & Tversky, A. 1982, "The Psychology of Preferences". *Scientific American*, 246, 136-142.
- Mandeville, B. de 1729, *The Fable of the Bees: or Private Vices, Publick Benefits*. Indianapolis (IN): Liberty Classics, 1988 (2 vols.).
- Newell, A. 1994, *Unified Theories of Cognition*. Cambridge (MA): The Harvard University Press.
- Russell, B. 1914, 'On Scientific Method in Philosophy'. In *The Collected Papers of Bertrand Russell* (ed. By J. Slater) – vol. 8. London: Allen & Unwin, 1986.
- Savage, L. J. 1954, *The Foundations of Statistics*. New York: Wiley & Sons.
- Schuck-Paim, C. & Kacelnik, A. 2002, "Rationality in risk-sensitive foraging choices by starlings". *Animal Behaviour*, 64, 869-879.
- Stanovich, K.E. 2013, "Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice". *Thinking & Reasoning*, 19, 1-26.
- Van Inwagen, P. 1983, *An Essay on Free Will*. Oxford: Clarendon Press.
- Von Neumann, J. & Morgenstern, O. 1944, *Theory of Games and Economic Behavior*. Princeton (NJ): Princeton University Press.
- Tversky, A. 1969, "Intransitivity of Preferences". *Psychological Review*, 76, 31-48.
- Tversky, A. 1975, "A critique of expected utility theory: Descriptive and normative considerations". *Erkenntnis*, 9, 163-173.
- Tversky, A. & Kahneman, D. 1981, "The framing of decisions and the psychology of choice". *Science*, 211, 453-458.