

DOI: 10.26319/EIDOS-001-FREEDOM

Marcin Poręba
Institute of Philosophy
University of Warsaw

Freedom, Symmetry Breaking and Reflective Judgements. An Attempt at an Incompatibilist Account of Freedom¹

Abstract:

In this paper free volitions are construed as a subclass of reflective judgements in the Kantian meaning, i.e. judgements not involving any fixed concepts but displaying a concept-like form. Judgements expressing volitions and issuing in action may be termed volitional judgements, therefore free volitions are construed as reflective volitional judgements. Due to an element of conceptual novelty and the fact that in a reflective judgement volition gets conceptually fixed out of an unstable equilibrium between different conceptualizations, free volitions thus construed seem to be apt candidates for causally undetermined events.

Keywords:

determinism, free will, freedom, incompatibilism, Kant, reflective judgement.

The paper presents a version of incompatibilism with respect to the free will problem. Free willings are construed as causally undetermined events, i.e. events not necessitated by anything in their past. However, on the account presented, there might exist a causal story accounting for the space of possibilities from which a free decision makes a choice. But since I understand freedom as involving an essential element of novelty, as bringing in

1) The author wishes to express his deep gratitude to a member of the editorial board of *Eidos* and two anonymous referees for their penetrating comments and suggestions which helped me considerably improve on the initial version of this paper. For all the inevitably remaining flaws, the responsibility is mine.

new elements into the network of already existing factors of human action, I strongly doubt if it could be made compatible even with weaker versions of causation, such as the nondeterministic (statistical, quantum) causation. However, much of what I claim in this paper might be considered indifferent with respect to these subtler questions as to which models of nondeterministic causation are consistent with the existence of free actions. My chief concern is rather with the question of what kinds of events picked out by some high-level characteristics, essential for our moral, legal and aesthetic discourse, in which the concept of freedom is rooted, could be identified with free events characterized metaphysically by the negative feature of not being deterministically caused – if there indeed are any such events. I propose to consider such high-level events willings, construed essentially as volitional judgements, i.e. items belonging to the class of conceptual operations, viewed basically after the Kantian model of judgement as a logical operation having concepts as its form. I propose to consider free volitions, as opposed to determined volitions e.g. by the needs and habits of agents, as pieces of reflective judgement, characterized by the absence of ready-made concepts, but displaying a concept-like form. As such they are equilibrium states between different possible conceptualizations, and the decision issuing from such a volition can be construed as a resolution of this equilibrium state in favor of one of the conceptualizations.

Let me begin with a couple of general remarks on the concept of freedom. It is certainly open to debate (and worth debating) to what extent our concept of freedom is a primitive one, i.e. independent of the concept of necessity, be it understood in causal or other terms. It is *prima facie* conceivable that free events exist in a world in which there is no necessity – no necessary causal laws and no necessary causal relations. In such a world of pure chance freedom would appear not as an exception to necessity but rather as an exception to pure chance and randomness. Certainly with respect to such a world the compatibilism vs. incompatibilism debate would take a different, perhaps more interesting form. It would then be a debate concerning the relationship between two primitive and equally fundamental concepts: pure chance (events undetermined by any causes nor causal laws) and freedom (events considered as intentional and as such falling under certain laws and constraints). And it is an interesting question whether these two notions of freedom (freedom as an exception to necessity and freedom as an exception to pure chance) are at bottom equivalent or not, and on what grounds. I even imagine that some important points in the account to be presented could be rephrased so as to be applicable to the second notion. But since the predominant form of the compatibilism debate concerns freedom understood as an exception to necessity, in the following I shall assume this notion as the concept of freedom.

Every philosophical theory of freedom has to meet at least the following two requirements:

- (a) it should decide whether exceptions to necessity are in the first place possible;
- (b) it should consider to what extent these exceptions correspond to our everyday, intuitive understanding of freedom.

There are at least three different notions of necessity and, respectively, three different notions of metaphysical freedom:

- (1) Logical necessity given by the metalogical principle of excluded middle: if every proposition at every point in time is determinately either true or false, then freedom is illusory insofar as it entails that at least some propositions concerning future events are undetermined as to their truth value.
- (2) Physical necessity understood in causal terms – philosophically the most widely used concept of necessity. Free events appear accordingly as exceptions to the causal closure of the physical domain.
- (3) Necessity as symmetry breaking – a concept remotely indebted to Leibniz' principle of sufficient reason, and lying at the heart of contemporary scientific theories. It refers to a situation in which

a system, being initially in an equilibrium state between several possible outcomes (symmetrical with respect to relevant physical laws) evolves towards one of these outcomes due to some minor fluctuations in the equilibrium state, so that from the outside the outcome may seem arbitrary.

Using (3) as a model for our considerations, freedom can be defined, respectively, as indeterminacy in symmetry breaking within a system whose behavior we typically describe as intentional. Crucial for this concept are obviously the kind of factors responsible for the indeterminacy. On the interpretation to be presented in this paper, these factors pertain to the conceptual nature of volition, construed after the Kantian model of reflective judgement. I elaborate a bit on this concept of freedom in order to show that it is in keeping with our intuitive notion thereof.

By “freedom” I shall understand what in Anglophone philosophy is typically meant by “free will” or “freedom of the will”. It refers to what we ordinarily have in mind when we speak of ourselves as “authors” of our deeds, when – looking into the past – we say “I could have done otherwise” or – thinking of the future – consider it essentially open insofar as it depends on our decisions. What we also have in mind is a particular feeling or sentiment – this is essentially how freedom appears at the moment of its actualization, when we freely decide to do something. Some of these moments or aspects of freedom can be challenged by dialectical arguments, most easily perhaps the moment of feeling – we all know how Spinoza reduced our feeling of freedom to mere self-consciousness: give a stone self-consciousness and it will immediately start to attribute its falling to its free decision.² Such a reduction or dissolution of all or some of these aspects is of course no theory of freedom, just as saying (for example) that there are no witches, and the putative cases of witchcraft are reducible to non-conventional healing practices, rather than a ‘theory’ of witchcraft. So what I am going to outline makes sense only provided that the majority of the attempted reductions with respect to the ordinary idea of freedom should ultimately prove to be ineffective.

One terminological point will be perhaps in order. The wisdom of the English usage “free will” instead of “freedom” lies in the fact that it blocks one typical line of reasoning that leads to a dissolution of the idea of freedom itself. Namely, if someone says, with reference to a situation in the past, “I could have done otherwise”, one might respond: “Yes, indeed. You could have done otherwise, if you had willed otherwise, and that’s precisely what it means that you are free”. And since this is perfectly compatible with a version of determinism on which our doings are “necessary *ex hypothesi*” (a phrase frequently used by Leibniz in this context), it follows that virtually every account of free agency has to be an account of free will, since it is only free will – if there is such a thing – that enters into the said *hypothesis* in a way that itself is not *ex hypothesi* necessary.

As far as the relationship between free will and causality is concerned, I consider incompatibilism to be altogether more plausible than its opposite. I am not convinced by any solution implying that one and the same event can be considered from two different standpoints, as on the one hand (deterministically or statistically) necessitated, and not necessitated (or free) on the other. And insofar as this is sometimes claimed to be an adequate interpretation of Kant’s account of free will, I’m against this account as well. But notice that at some crucial points of his discussion of free will, Kant uses as one of his premises the implication from “I should (in the moral sense) have done otherwise” to “I could have done otherwise”, having obviously in mind that even if all circumstances of my decision had been kept unchanged, I still could have decided otherwise.³ I take this

2) See e.g. Letter LVII, in: Benedict de Spinoza, *The Chief Works of Benedict de Spinoza* (London: George Bell and Sons, 1901) Vol. 2, 390–391.

3) Immanuel Kant, *Critique of Practical Reason*, Werner S. Pluhar trans. (Indianapolis: Hackett Publishing Company, 2002), 4–5, 42–44.

as a strong suggestion that he was also an incompatibilist, and that the causal closure of the physical domain of which he speaks (e.g. in the second of the “Analogies of Experience”)⁴ has to be interpreted differently than it is typically understood.

My chief reason for rejecting compatibilism lies in the fact that one of the aspects of our idea of causality is to narrow down the scope of what is possible in the future of any given point in space-time, where “future” denotes the contents of the forward light cone of the event in question, i.e. all the events that can be causally influenced by the event in question. Determinism implies that only one such future is possible. To be sure, there are also other aspects to the idea of causality, due to which determinism – even the strictest one – is different from fatalism. But certainly determinism, insofar as it implies causal closure, agrees with fatalism on the question of whether the future of any given point could be different. This in turn can be explained by the fact that both fatalism and different versions of determinism are just more or less philosophically sophisticated attempts to justify the idea of reality as “what is out there”, “out there” meaning “not here”, not where we are. This “not here”, or “outside” can be understood both in a spatial and temporal sense. The most frequently discussed idea of reality of what is “outside the mind” is only a special case of this general concept of reality as `what is out there`.

There is a close connection between the idea of reality as something fixed and determined and the idea of there being a kind of mechanism securing this fixed status of reality, so to speak “tracking” (be this mechanism brute and unintelligible as in fatalism or so to speak rational as in causal determinism). It can be clearly seen not only from the etymology of the word “to determine” (Latin *determinare*) but also from the fact that, on some influential accounts of indeterminism, it implies some form of anti-realism with respect to the future (or the past, for that matter). This kind of anti-realism is perhaps most famously exemplified by the Polish logician and philosopher Jan Łukasiewicz⁵ and the eminent English philosopher Michael Dummett.⁶ Both claimed it necessary to suspend a portion of classical logic in order to save our intuitive conviction that the reality “out there” is not determined (causally or in some other way) by what is “here and now”. Such a policy is supported by the implication that if the reality out there is logically fixed (if every well-formed proposition about it is “determinately either true or false”, in other words, if it obeys the metalogical principle of excluded middle), then it is also fixed in some extra-logical way.

However, I reject this implication, because I believe that we can make perfect sense of the idea of a fixed reality without falling back on some extra-logical reality-fixing mechanism. But the foregoing consideration shows some aspect of determinism which allows for a better understanding of my rejection of compatibilism. Namely, if I am right in believing that the motivation behind determinism is primarily to secure our idea of a fixed reality and only secondarily to account for the causal makeup of the world, then determinism seems to be an unnecessarily strong assumption. Even if it is true, with respect to the future, that p or q , it by no means follows that either necessarily p or necessarily q . The appearance to the contrary stems from a conflation of two distribution principles: $\Box(A \vee B) \rightarrow (\Box A \vee \Box B)$ (valid in so-called normal modal logics) and the respective distribution principle of temporal logic $F(A \vee B) \rightarrow (FA \vee FB)$, where F means “it will be the case that”. Since

4) Immanuel Kant, *Critique of Pure Reason*, Paul Guyer and Allen W. Wood trans. (Cambridge: Cambridge University Press, 1998) 304 ff.

5) See e.g. his *On Three-Valued Logic*, in: *Selected Works*, ed. by L. Borkowski, (Amsterdam/London: North-Holland Publishing Company, 1970), 85–86; *On Determinism*, *ibid.*, 110–128; *Philosophical Remarks on Many-Valued Systems of Propositional Logic*, *ibid.*, 53–178 (esp. 176–178).

6) Michael Dummett, *The Reality of the Past*, in: *Truth and Other Enigmas*, London: Duckworth, 1978), 358–374; *Tense and Time*, in: *Thought and Reality* (Oxford: Clarendon Press, 2006), 73–84.

from purely formal and algebraic point of view both principles behave very similarly, it is tempting to interpret them as two ways of expressing one and the same, very general truth that whenever it is fixed that either p or q , it is either fixed that p or that q . On this interpretation it seems not only harmless but even revealing to say that whenever it is fixed that either p or q , it is either necessarily fixed that p or necessarily fixed that q . To embrace this would clearly amount to a sort of determinism incompatible with the idea of free will: even if we are ignorant as to which of the disjuncts will turn out true, it will turn out true necessarily, and our idea that we have free choice reduces to our ignorance of the outcome and thus dissolves. This is why I am an incompatibilist, but my rejection of determinism is not so much motivated by my belief in human freedom, but rather by the more general considerations (as sketched above) concerning the roots of deterministic thinking, no matter what bearing it has on the problem of free will.

Rejection of determinism means that at least for some points in space-time what happens at them is not necessitated by their respective pasts. This by no means entails the denial of a causal structure of the universe. It could be even argued that some such structure is necessary in order to secure the distinction of past and future, and if we do not want to admit of points having no definite spatiotemporal location, it seems even to entail that at every point in space-time there is (or happens, or is instantiated) something which causally depends on the past of this point. In light of this remark, indeterminism seems to be the thesis that for some points in space-time part of what happens at them is *not* determined causally by their past.

What is needed to build indeterminism thus defined into a theory of human freedom are the following two points:

- (1) At least some of the events undetermined by their past should be identified with operations of free will.
- (2) At least some of the events undetermined by their past – notably the ones identifiable with free willings – should have causal bearing on their future.

Point (1) is clearly more problematic since it entails that a subclass of a very abstractly defined class of events, having in common just one ontologically primordial character (“events causally undetermined by their past”) systematically coincides with a set of events picked out by some extraordinarily high-level characteristics, deeply rooted in our everyday experience and including portions of religious, philosophical and legal vocabulary. The fairly uncontroversial examples of causally undetermined events stem from the quantum level – it seems, for example, undetermined which one of the atoms of some radioactive element will decay at a given moment; however, it would be extremely odd to say that it is a matter of free choice on the part of the atom or of the sample to which it belongs. And I think it rather implausible to say that the indeterminacy of free will is at bottom quantum indeterminacy – after all it would contradict some of the most unshakable principles of quantum physics such as complementarity, since the exercises of free will clearly belong to the “classical” part of reality, i.e. to the part which is at least graspable without appeal to quantum indeterminacy. Therefore if we want a theory of free will which naturally extends our intuitive understanding thereof, we need first some positive account of causal indeterminacy, independent of its quantum exemplifications.

As the starting point I propose to consider Leibniz’ principle of sufficient reason. It is frequently misunderstood as a form of causality principle, whereas in fact causality is what operates *w i t h i n* the actual (or any other possible) world and *n o t* at the level of choosing between possible worlds, to which the principle of reason applies. With the help of the principle of sufficient reason, Leibniz answers his famous question “Why is there something rather than nothing?”, where “nothing” is to be understood not in the Parmenidean sense of emptiness or less than something but rather in the sense of *m o r e* than something, in the sense of the

entire space of possibilities as opposed to one of its points (or regions)⁷ which happens to be the actual world. The choice between these possibilities, issuing in the actual world, is a clear case of symmetry breaking: all possible worlds, construed as ideas in God's mind, are perfectly symmetrical with respect to all laws of logic and metaphysics, and what decides each outcome are some tiny differences in the distribution of qualities between them. More than that, it is rather a case of spontaneous symmetry breaking than of an explicit one,⁸ since it happens not due to the operation of some causal law disturbing the initial symmetry but rather due to a free decision of God.

Kant applied this spontaneous symmetry-breaking model at several points in his philosophy, most importantly perhaps to the question of the relationship of appearances to things in themselves and to the closely related question of free will. There are two fundamental misunderstandings made with respect to this question. First, Kant's freedom ("transcendental freedom" as he sometimes calls it) is frequently construed as depending on moral law, i.e. only as acting morally can I be free. In fact, moral law which unconditionally demands an action only makes us aware of our freedom – it by no means constitutes it. Second, Kant's solution of the third antinomy of pure reason, i.e. the antinomy of strict determinism and freedom, is sometimes read as implying that both terms can retain their full validity, since determinism refers to human agency considered as appearances in space and time, and freedom to the same agency as a thing in itself.

In fact the Kantian solution is much more sophisticated and requires a fundamental reconsideration of both terms. What remains of determinism is the second analogy of experience, interpreted as a synthetic *a priori* principle forbidding us to think of any particular event in space and time that it has no cause. But this principle by no means precludes thinking that there are events not causally necessitated; it only says that it is forever beyond our cognitive powers to decide of any particular event that it has this character of causal indeterminacy. So Kant's determinism is by no means strict. Causal determination pertains only to possible experience, and even this is only insofar as part of our possible experience is our conceptual disposition to think in terms of causes and effects. In other words, insofar as "possible experience" means "empirically justified true beliefs about objects and events in space and time", causally undetermined events are indeed beyond experience, and as such a "thing in itself". But insofar as possible experience is understood as the domain of objects and events to which our empirical beliefs refer, we are not merely entitled, but obliged to assume the existence of causally undetermined events, since only on this assumption does moral law make sense.⁹

Now let us consider events which, however they are parts of possible experience, are not causally necessitated by other parts thereof. Do they constitute as acts of free will? It is rather implausible, considering the fact that our concept of free will is strongly embedded in moral and legal discourse, and the concept of a causally undetermined event is a purely metaphysical one. Moral considerations assure us that there are such events, but they by no means lead to the conclusion that the only causally undetermined events are the morally relevant ones. Free decisions use, so to speak, the model of causally undetermined happenings which cover a much broader class of events. What is the *differentia* of freedom with respect to the possibly broader class of the causally undetermined? I can see no essential difference here. I prefer to think of it rather in terms of characteristics

7) Depending on whether "possibility" is taken to mean a set of jointly possible states of affairs or a single such state of affairs.

8) This important distinction refers to the way relevant laws contribute to the evolution of a given symmetry state. Explicit symmetry breaking is where the outcome is somehow determined by a law, spontaneous symmetry breaking is where the outcome is determined rather by the structure of space within which a given system evolves. In the Leibnizian scenario this is the space of all possibilities and the distribution of perfection between possible worlds is clearly a feature of this space not implied by any of the laws relevant for God's decisions.

9) A detailed interpretation of Kant's conception of causality along these lines is developed in Marcin Poręba, *Możliwość rozumu. Ćwiczenia z metafizyki* (*The Possibility of Reason. Exercises in Metaphysics*), (Warszawa: Aletheia, 2008), 301–309 and 380–386.

that some causally undetermined events happen to have. One of them is the fact that they issue in human agency. The other is that they play a very specific role in moral discourse. But none of these two makes them what they are metaphysically: gaps in the causal order of the universe.

A chance of linking together the two principal aspects of the concept of freedom – causal indeterminacy and agency under moral or similar constraints – lies in clarifying the account of what the aforementioned indeterminacy amounts to. Causal indeterminacy makes sense only against the background of a network of causal processes. The same is true of the concept of moral agency. To be able to say “I should (and therefore could) have done otherwise”, expressly referring to the past, I have (as a matter of a transcendental possibility condition) to be able to locate some events in my past instead of my future, and this in turn is possible only by reference to a causal story about my life and its place in a broader world-scenario. To say that a certain event in the past was an exercise of free will means, metaphysically speaking, that it was (1) causally undetermined and (2) located in the spatiotemporal vicinity of some causally determined events (as such events typically feature publically observable pieces of behavior), since otherwise it would be explanatorily empty and there would be no reason to postulate free decisions in the first place. (1) is a purely negative point, whereas (2) is perfectly true of *any* event belonging to the past. So we clearly need more.

What I propose in the following is based on two considerations. First, it seems useful for our purposes to approach the idea of causal indeterminacy not from the abstract and purely negative perspective of an event not being caused (necessitated) by anything in its past, but rather from the more specific standpoint of an event being a case of spontaneous symmetry breaking. Second, since spontaneous symmetry breaking takes place both in intentional and non-intentional systems, we need some differentia of the intentional ones and among them in particular of moral agents. As such I propose the conceptual character of the relevant symmetry breakings. By their conceptual character I mean that they can be interpreted as exercises of concepts.

Spontaneous symmetry breaking, as opposed to explicit symmetry breaking,¹⁰ means that there is no lawlike causal story about why it occurred and why it took precisely such and such shape. It can be said only that the symmetry to be broken has to be unstable, as for example in the case of a radioactive atom. Applying this to human agency it can be said that one cannot stay for very long in a state of indecisiveness, but has to decide some way. However this “has to decide” is frequently given moral or existential interpretation, it boils down metaphysically to the decision’s being necessitated by the instability of a symmetry state whose resolution it is. But unlike the decision as such, its precise content is not determined by the necessity of leaving the state of indecisiveness.

But are there in the first place any such states of symmetry in human decisions, issuing in spontaneous symmetry breaking? If we are to be considered free agents, then certainly there are, but are there any independent reasons for believing that there are such states? I am aware of two such reasons. The first is the observation, that in systems such as intentional agents, states of instable equilibrium (symmetry) abound and it is *a priori* rather implausible that the way symmetry gets broken in them is in every case explicitly causally determined. Second, there is the understanding, first conceived by Kant, about how some such states can be construed as states of rational agents, providing an independent argument in favor of the non-causal character of their evolution. Let me elaborate a bit on this Kantian proposal.

For Kant all intentional states, no matter if they belong to cognition or action, are at bottom judgements, i.e. exercises of concepts. Judgements can be of two kinds: determining and reflective. The difference between both kinds lies in the way concepts are involved in them. Determining judgements are exercises of

10) See above, ft. 8.

concepts already possessed and registered; in reflective judgements a new conceptual function comes into play. Kant describes reflective judgements (especially aesthetic ones) explicitly as equilibrium states, as a “free play” of cognitive powers, in which concept-like regularities occur and which come to an end once one such regularity gets fixed and thus becomes a concept.¹¹ Kant’s analysis of this situation, originally meant as an explanation of reflective judgements about the beautiful, admits in my opinion of a much broader class of applications, including both cognitive and practical judgements. In the cognitive domain it has the potential of accounting for the possibility of conceptual change, which offers the practical function of explaining exercises of free will.

The following objection can be raised against the last kind of application. Practical judgements (judgements expressing moral duty or “the ought”) are considered by Kant as typical cases of determining judgements, based on the concept of moral law (categorical imperative). So it seems highly implausible that these judgements should involve an element of reflection as well. On the other hand, Kant was acutely aware of the gap between general statements of moral obligation and particular volitions and decisions issuing therefrom. The crucial point in Kant’s discussion of moral decisions is that the application of the generalizability criterion (“Act only in accordance with that maxim through which you can at the same time will that it become a universal law.”)¹² to a particular morally relevant situation requires finding a maxim adequately expressing a particular volition whose moral admissibility one wants to probe. It is true that once such a maxim is given it is a matter of a determining judgement to generalize it to a universal law, which allows for an application of the criterion. But what is *not* a matter of a determining judgement is the maxim itself, construed as an essentially conceptual achievement, as a judgement which gives my volition its form, or better which constitutes my volition, since there is no such thing as a ready-made volition which only has to take the form of a judgement, but forming a judgement is rather an essential part of making the volition itself.

Forming a judgement (volitional judgements being no exception in this respect) is an essentially conceptual achievement, and as such a matter of what Kant calls “spontaneity”. Generally speaking, spontaneity means that however much a judgement is always a response to some challenge or demand (be it an array of sensory stimuli, a felt tension in one’s belief system or a practical need), its precise content is never determined by this triggering condition (can never be deduced therefrom), but depends on the subject’s conceptual resources. In determining judgements it depends on some well-established concepts (on Goodman’s model of induction, on some generalizations well-entrenched through repeated projections in the past),¹³ in reflective ones it depends on a more general faculty of conceptual thinking not explicitly “programmed” by any concepts, but capable of self-programming. So it seems that reflective judgements are marked by a twofold spontaneity: a spontaneity to the second power, since they are neither determined by the situation to which they respond nor by any concepts already possessed by the subject.

Now let me express my main contention in this paper: if there are judgements that are on the one hand reflective (spontaneous in the strong sense of being determined neither by their triggering conditions nor by any concepts preceding them) and on the other volitional (issuing in intentional action), then these judgements, or better *judgings*,¹⁴ considered as particular events, are good candidates for free willings. The rest of my paper will be a brief comment on this contention.

11) Immanuel Kant, *Critique of the Power of Judgement*, Paul Guyer and Eric Matthews trans. and eds. (Cambridge: Cambridge University Press, 2000), 102–104 and 124–127.

12) Immanuel Kant, *Groundwork for the Metaphysics of Morals*, Allen W. Wood trans. (Connecticut: Yale University Press 2002), 37.

13) See Nelson Goodman, *Fact, Fiction, and Forecast*, 4th edition (Massachusetts: Harvard University Press, 1983), 84–99.

14) By “judgings” I mean particular performances of logical operations whose contents are respective judgements.

First of all, the observation that forming a volition might be a conceptual achievement of the reflective sort seems to be backed by everyday experience of coming to terms with nonstandard situations (situations that do not clearly fit any but some very general pattern or in which our desires and inclinations are unclear to us). Part of coming to terms with such situations can be a conceptualization which cannot be fully traced back to any concept already possessed. Surely not every situation is of this kind, but on the other hand, do we not imply that a situation was of this kind, when we say with respect to it: “I could have done otherwise”? What makes us do wrong is mostly not bad will nor *akrasia*, but rather wrong conceptualization where some good (be it egoistic or altruistic) is considered as a morally admissible reason for action. “Could have done otherwise” boils this way down to “could have judged otherwise”. Note that this by no means implies ethical intellectualism, since (1) the good considered as a moral reason might be egoistic, altruistic or even cosmic (i.e. the perfection of the Universe) and (2) it is precisely the indeterminacy of volitional judgements by any conceptual considerations that constitutes the essence of morally relevant situations, whereas according to ethical intellectualism there is always a conceptually sufficient condition of an action’s being morally good.

Second, even though conceptual indeterminacy of a volitional judgement does not logically imply its causal indeterminacy, the judgement has the symmetry breaking structure desired for free events. A typical symmetry in such cases is between a conceptualization in strictly moral terms (in terms of the unconditional ought) and in terms of some presumed good outcome. Whether the issuing judgement is in fact causally undetermined is certainly a question not to be decided on purely philosophical grounds. But if there are any causally undetermined events at all in the order of nature, reflective volitional judgements seem to be apposite candidates to claim this status.

Third, as an additional comment to this account of free choice as issuing from a reflective volitional judgement, it will be perhaps in order to compare it with the account on which “free choice” is a primitive concept and the conceptualization given frequently as a reason for action is at best a rationalization, if not straightforwardly bad faith.¹⁵ This type of account is best exemplified by existentialist theories of human agency, most notably those of M. Heidegger and J.-P. Sartre. I think that the key to understanding the relationship between the two accounts lies in the notion of reflexivity of volitional judgements with which I propose to identify free decisions. Crucial for this notion is the idea that a conceptualization issuing from a reflective judgement always involves an essentially new conceptual element, whereas in the case of rationalization or bad faith we have always to do with an attempt to explain the choice in terms of conceptual resources already possessed at the moment of entering the decision-making process. It seems therefore that the account outlined in this paper could be given an existentialist interpretation (however, I do not think that it is the only possible interpretation thereof), so there is no obvious incompatibility between the two accounts. On the other hand, it seems to me that the notion of reflexivity as involving an element of conceptual innovation could be very useful in the context of an existentialist philosophy of mind, insofar as new concepts, as involving always a new decision concerning how to think, would be never deducible from the old ones. If this is correct, then the existentialist theories of human agency could be interpreted as a special case of the reflective theory sketched above: they imply it but are not in turn implied by it.

15) I owe the insight into the need of clarifying this issue to one of the anonymous referees of this paper.