

16

The Core of Free Will

Wolfgang Spohn

Department of Philosophy, University of Konstanz

Everyone sober believes in freedom of the will. Whatever we precisely mean by it, it is something we have. The world, which includes us, may be deterministic. This generates a contradiction. Or the world may be indeterministic, but it seems common ground that this does not improve the dialectical situation; the freely willed cannot occur at random. There is no escape; we are dealing here with a sharp contradiction, an outright antinomy.

When I look at the literature on the problem of free will—I am sure the majority has escaped me—I am amazed at how many different thoughts and theories are provoked by it, each being relevant, well considered, a germ of a larger research program, and containing a grain of truth (or more). This variety of responses is what makes for a truly deep and fruitful philosophical problem. Here, my point will not be to offer any novel thoughts on the topic; every stone has been turned many times. Despite its richness, however, the literature appears to me to give only improper emphasis to a certain point that occurred to me in writing my dissertation (1978, 193) and which seems ever more central to me, the longer I observe the literature. Mathematicians know that a sharp contradiction needs a sharp answer; wiggling concepts till the contradiction dissolves is not allowed.¹ Let me explain what I take to be a sharp answer.

I

The rock-bottom fact is that there is a normative point of view and an empirical point of view. This is trivial and uncontroversial. The point I am going to explain, though, will be that even as an empirical scientist, you cannot get rid of or even delimit the normative point of view. It spreads to all human affairs, and you cannot empirically investigate them without yourself engaging with the normative point of view.

What has this to do with the freedom of the will? The answer is immediate. For, what is the normative point of view? It consists in asking normative questions and seeking normative answers. The paradigmatic normative question is: "What shall I do?" By contrast, the paradigmatic empirical question is: "What has happened?" or "What will happen?" And more specifically concerning human affairs: "What will he do?" or "What will she do?"

I seem to be identifying the normative with the first person or the subject's perspective, and the empirical with the third person or the observer's perspective. In a way, I do. Of course, I grant that the distinctions do not coincide. I can think about normative questions concerning other persons, and I can consider empirical questions about me; I can even try to predict future actions of mine. Still, the normative force of an answer to a normative question is immediate only in the first-person perspective. My answer to another person's normative question develops normative force only when the other person poses this question to herself and accepts my answer. In this sense, the normative is peculiar to the first-person or subject perspective and is never to be gained from the third-person or observer perspective alone.

Some have a narrower understanding of the normative. Normative issues are concerned with norms, and norms (or maxims or rules) are for everyone. Maybe. However, we need not worry here about the generalizability of answers to normative questions. Others prefer to distinguish the prescriptive and the descriptive. I take this to be the same as the distinction between the normative and the empirical.

Note that the distinction between normative and empirical questions is not the same as that between practical and theoretical issues. Theoretical issues are ambivalent as well. I may ask: "What is the case?" Or I may ask: "What shall I believe to be the case?" In a way, I am asking here the same question twice, once in the observer's mode and once in the subject's mode. Hence, the distinction between the

normative and the empirical cuts across the theoretical and the practical. However, since I do not want to reflect on normative epistemology, I restrict myself to the practical side.

Now, it is impossible for us to avoid the normative question: "What shall I do?" Even refusing to answer it is an answer, a decision to let things go. Often it happens to us that we unintentionally ignore the question, because we cannot always keep everything under control. (It is even nearly impossible for us to keep track of everything that is in principle under our control.) However, we cannot unintentionally ignore the normative question all the time. (At best, though I doubt it, we might attempt to reach such a state of entirely banning the normative perspective, which some may call true nirvana.)

Since we cannot avoid the normative question: "What shall I do?" we must ask next: "How do we go about answering it?" How does practical deliberation work? This is indeed a complicated matter. Undeniably, though, our most instructive and powerful abstract model of practical deliberation is decision theory. So, let us look at it. I have argued (1977, 114ff.; 1978, secs. 2.5 and 5.2) for the so-called *no probabilities for acts* principle, which says that a decision model must not contain the subjective probabilities of the agent for his possible actions that are at issue in the model. In more ordinary terms, the point is simply this: When you try to answer the question: "What shall I do? A or B or C?" you do not have an epistemic attitude toward the possible answers A, B, or C. Rather, you reflect on your aims and values (which are also open to normative discussion, of course), you consider your relevant beliefs about the world, about the consequences of your possible actions, about the attainability of your aims, and so on, and from all this you try to get at a normative conclusion about which option is best. In all this, no epistemic assessment of your possible actions is involved. The question of how likely you are to do A, B, or C is simply not part of your practical deliberation.²

The next point is that this *no probabilities for acts* principle entails the *acts are exogenous* principle, which says that the possible actions at issue in a decision model are exogenous, that is, *first causes* or *uncaused* within this model. Of course, this inference is, and needs to be, backed up by a sophisticated account of causation.³ However, the point is also intuitively clear without that account. It is simply that as soon as you start thinking about the possible causes of your actions, you are about to epistemically assess your possible actions, you are leaving the sub-

ject's normative perspective and adopting the observer's empirical one. This is not to deny determinism in order to accept indeterminism. Indeterminism does not defy epistemic assessment; on the contrary, it only makes it ineliminably probabilistic. The point rather is that the normative perspective sidesteps the determinism/indeterminism issue altogether, as we know the defender of the freedom of the will must do.

In a way, one must admire Kant for his uncompromising attitude. He accepts freedom of the will, he accepts determinism, *and* he accepts the contradiction. He is entirely clear that only a radical solution will do, and he finds one in his two-worlds doctrine: determinism holds in the phenomenal world, freedom is in the noumenal world, and thus the contradiction vanishes (see Kant 1785, sec. 3, in particular pp. 76ff.). Of course, this doctrine is absurd; there is only one world (and even a thoroughgoing argument with Kant could not come to a different conclusion). But it brilliantly shows what is at stake in a principled solution.

What I would like to convey is that the distinction between the normative and the empirical perspectives has the same principled force. There is only one world, but two perspectives. In the subject's normative perspective, his own actions are uncaused, and this is compatible with the fact that they may be fully determined, that is, sufficiently caused (or only partially determined, that is, probabilistically caused) in the external observer's empirical perspective. This is what I take to be the core of free will. If you search for it only within the empirical perspective, you are lost in paradox; within the normative perspective its existence is a truism.

To connect up with the received terminology: I plead for compatibilism, but only by distinguishing two perspectives. There is no compatibilism within each perspective; the subject's normative perspective is clearly libertarian, whereas the observer's empirical perspective may well accept full determinism.

Perspective-dependent causation: what a strange idea, you will object. And surely, the perspectives cannot be on a par. Before attending to such worries, let me pause to mention that my point is a very old idea; as I said, every stone has been turned many times. Obviously, there is something special about the causation of our actions. The causal mystery is not on the cosmological scale of a first mover or a *causa sui*, but on a smaller, human scale. At least since Thomas Reid, the idea of agent causality, as it was called later, has been with us: my

actions, or volitions, are certainly caused, but they do not simply fall into the overall causal nexus of events. Rather, *I* cause them, *I* do or will them. Some philosophers find this idea natural; most puzzle over it. Perhaps it can be better understood by distinguishing different perspectives. Within a single perspective, as it is usually presented, it is bound to remain mysterious.

The idea that one should therefore distinguish various perspectives (worlds?) is also very old. For instance, Daniel Dennett (1984) exploits his distinction between the physical stance, the design stance, the intentional stance, and the personal stance to account for free will. Thomas Nagel (1986, chap. 7) clearly argues that freedom, autonomy, and responsibility look different from the subjective and the objective perspectives. The point took on a more dramatic dimension with Georg Henrik von Wright (1971), who, by opposing explanation and understanding, indeed argued for a principled and unbridgeable difference between the sciences and the humanities; this is, of course, a topic with a long history. Julian Nida-Rümelin (2005, in particular essay 2) ~~chimes in with his thesis of the nonnaturalizability of reasons and the complementarity of a naturalistic and a humanistic perspective.~~⁴

These references may suffice. Why I still feel urged to write this note is merely a matter of sharpening. Somehow, I sense only insufficient awareness of the fact, outlined above, that the causal relations are to be judged differently in the two perspectives, that is, of the fact that the distinction between the two perspectives and the uncausedness of actions within the subject's normative perspective is a necessary and rigorous consequence of the received view of practical deliberation and a most prominent view of causation and causal dependence. Neither do I find proper emphasis in my immediate allies: Judea Pearl (2000) and Peter Spirtes et al. (1993) share my premises, but they do not relate them to the freedom of the will, whereas Isaac Levi (1986, 65–66), who relates practical deliberation and freedom in the way indicated, remains silent on causation.⁵

II

Still, my outline above would be badly incomplete if I left the relation between the two perspectives without comment. The first worry is, of course, that I am obviously assuming perspective-dependent causation and that this is hardly better than Kant's two-worlds doctrine. How-

ever, the situation is not so bad. There is a precise account of how the causal pictures relate according to the two perspectives. We may represent the causal relations according to the empirical perspective by a causal graph, and then we arrive at the causal relations according to the subject's normative perspective by truncating, as it is called, the empirical causal graph with respect to the action nodes. Thus, the causal relations mostly agree. No less and no more than the causal relations that have the subject's actions as immediate effects are cut out; the actions are, as it were, the causal blind spots of the normative perspective.⁶

However, do I not grant thereby that it is the empirical perspective that offers the real, complete causal perspective filling the local blind spots of the normative perspective? Yes, I do. Therefore, should I also grant that the empirical perspective is the primary one, the only one that counts in the end? And then are we stuck in our paradox as badly as ever? No, I deny both. Neither of the two perspectives is primary; they are on a par and have a rather complicated relation. Let me explain this point in more detail.

First, there is no fundamental mystery whatsoever for the observer's empirical perspective as to how actions are caused. Ideally, it is precisely the practical deliberation that causes the ensuing action. This formulation suggests, though, that we actually deliberate all the time. This should not be implied. We should rather say that the action is caused by the mental configuration represented by the decision model appropriate for the case at hand.⁷

Of course, decision theory only provides the basic pattern. Even if this pattern were roughly true, we should inquire into how this precisely works, what the underlying mechanisms are, how that mental configuration of beliefs and desires is caused in turn, and so on. These inquiries should proceed on a psychological as well as a neurophysiological level. This is already an inexhaustible research agenda. There is, moreover, the well-known concern that decision theory is not even nearly empirically correct, and thereby empirical research gets even more demanding and involved.

However, I am not so impressed by this concern. Decision theory is a normative theory, in the first place, designed for the normative perspective. Although one need not deny the relevance of empirical findings for normative discussion, which normative conclusions to draw from

them is quite unclear (see Spohn 1993). Hence, I think that the criticism of decision theory from this side is at least short-circuited.

I am more concerned by the fact that, even as a normative account, decision theory is defective. I am not alluding to various normative doubts about the basic principle of maximizing (conditional) expected utility, which indeed are more or less pressing. The point is rather that the intrinsic (nonexpected) utility function is treated simply as given in decision models, although it can and should be subject to normative discussion in a number of ways. Indeed, this is the place where I locate most of the literature on free will, even if it is not phrased in decision theoretic terms. The libertarians deny determinism outright, the hard determinists deny free will, and those like me distinguish two worlds, perspectives, and so on, to make things compatible; all three groups are minorities. The large majority seeks to realize compatibilism by saying that an action is free if and only if it is *appropriately* caused, and then all effort goes into explicating what “appropriately” is to mean here.

In order to be free, an action must be an action, not merely behavior, and it must not be coerced or compulsory. It must not merely satisfy fixed and given desires or utility functions. It must rather be responsive to reason in a more comprehensive sense of “reason” than the merely instrumental one (whatever this sense may be). The first-order desires must in turn be responsive to or under control of certain second-order desires. Maybe, they have to survive cognitive psychotherapy. Such views may also be related to suitable senses of autonomy. The subject must have had the opportunity to develop his or her *own* aims or desires in a sufficiently self-determined and reflective way, or she must take a stance toward them and accept them as her own. Even more directly, the (first-order) desires must have the right kind of content. They must conform to moral duty (and the categorical imperative), or they must be humanly adequate in respecting our rational nature or in perfecting our virtues. These and other ideas are discussed with great care in the philosophical literature. In all this, one must never forget that an appropriate notion of freedom goes hand in hand with appropriate notions of human dignity, responsibility, and blameworthiness with all its practical, moral, and legal implications.

The allusions in the preceding paragraph are too numerous to make all of them explicit. Also, it is not necessary to do so, since I am not going to discuss any of these views. I only want to point out that there

is a very rich, interesting, and relevant body of literature in some way or other occupied with completing our basic causal explanation of actions. Note, moreover, that the manner of argument is usually normative rather than empirical. The claims are that our actions should be guided by moral motives, they should be responsive to reason in a more comprehensive sense, they should be governed by second-order desires, and so forth. The extent to which they really are is not the philosophical issue, though the presumption certainly is that normative considerations have a real impact.

Still, all this does not detract from the fact that the basic explanation is the decision-theoretic one I have indicated above, however it is to be amended. Indeed, it is clearly the only explanation consistent with our self-understanding as beings necessarily endowed with a normative perspective.

So, of course, our actions are caused. They are mentally caused, as just outlined. They are even physically caused. I feel comfortable as a type-type identity theorist: mental states supervene on physical states (perhaps to be taken in some suitably wide sense); and depending on one's notion of a property or a state, this entails that mental states *are* (identical to) physical states. Likewise for causation. Causation exists on all levels, not only among elementary particles. There are causal relations among mental states (and between mental and physical states), and if mental states are physical states, these mental causal relations *are* physical causal relations; I do not see particular problems with supervenient causation.

These are apodictic claims stirring up a philosophical snake pit. Even to start defending them is far beyond the scope of this paper. However, it is not necessary to do so. The dialectics is rather that I think I can grant my imaginary opponent the strong materialist position of a type-type identity theory; I do not have to try my luck with some sophisticated doctrine about the relation between the mental and the physical that may open some tangled argument. I am even willing to grant that our normative point of view supervenes on our physical constitution. If our normative conceptions differed from what they presently are, the distribution of matter would have to differ, too, from what it presently is.

This is not to say, though, that the normative facts themselves supervene on physical facts. I do not know whether there are any. Perhaps

normative truths are those we would arrive at in the Peircean ideal limit of normative (not empirical) inquiry. However, one may well doubt that this limit is well defined. In any case, this limit, even if well defined, is highly counterfactual, and the (Humean) supervenience of the counterfactual on the factual is a more doubtful matter. Hence the materialistic supervenience of our actual normative opinions has no consequences for the supervenience of normative truths. We may, and should, leave open this issue.

The point I want to make now is that even the ontological professions of an identity theorist, which I share, by no means determine our epistemological or empirical third-person perspective. They do not decide the primacy of the empirical over the normative point of view, and they do not undermine the ineliminability of the normative point of view. Why is this so?

It is a well-known philosophical maneuver to turn ontological considerations into epistemological ones with the help of Laplace's demon. By knowing the ultimate ontological inventory of our world, the distribution of matter (at a given time) and the fundamental physical laws governing it, the demon can apparently know everything that is, and he can apparently explain every past and predict every future action and even every normative conception we will have. He seems to be the incarnation of our epistemologically perfected empirical perspective, and there is no place for the normative perspective in that perfection.

However, this is a seriously deceptive picture. We need to understand how wildly nonhuman the demon is. The point is not that in our indeterministic universe even the demon would not get far. Ontologically, we may well assume strict determinism. The point is rather that neither we nor the demon are capable of specifying the supervenience relation that is only claimed to obtain in our ontological professions, and that this incapability has very different, though converging reasons for us and the demon.

For us, the problem is not so much complete knowledge of fundamental physical laws; perhaps we are on the verge of it. For us, it is rather the demon's complete knowledge of particular physical facts (at a given time) and his perfect computational capacities. Both are entirely fictitious for us. It is safe to predict that we shall never exactly compute complex molecules in quantum mechanical terms and that, despite the bold declarations of neuroscientists, we can never have

more than the roughest understanding of the physiological supervenience base of mental processes like, for example, those involved in producing and grasping the sentence just written here.

In particular, we have to proceed from the above simple explanation of our actions, which was the only one consistent with our having a normative perspective. We may and should specify, qualify, and amend it in multifarious ways, some of which I have indicated above. Of course, we also evolve our normative point of view; we seek ever better and more complete answers to our normative questions. At the same time, we thereby promote our empirical perspective; our normative conception serves as well as our empirical ideal. We often do what we should, and we often do not; we often fail and often live up to our normative ideal, the various forms of failure are discussed philosophically under the heading of the weakness of the will. Every empirical theory about our behavior must respect this point by taking the normative theory as an idealization (just like, say, frictionless motion) and by complementing the ideal by various error theories. Any empirical theory that simply neglects our normative point of view is bound to be incomplete and inadequate.⁸

The demon has the complementary problem. Well, not necessarily, the demon might also be an eliminativist and thus not care about supervenience. However, I take the eliminativist's prediction that our intentional idiom will eventually dissolve to be simply incredible. Hence, if eliminativism is no option, then it will not do for the demon to know everything there is to know on the basic ontological level of physics. He is still entirely ignorant of all relevant supervenience relations. If he wants to know what water is, he must first know our notion of water; then, of course, it is easy for him to establish that water is H₂O. If he is to predict whether or not I am happy tomorrow he must also know how happiness supervenes on all the physics he knows; and in order to know this he would first have to acquire the complex notion of happiness. Likewise for all the other mental concepts we have. In particular, he would need to have and exercise a normative perspective by himself; otherwise, he could never grasp what our normative discourse is all about.

From both sides, we thus arrive at the same conclusion. The demon needs to have a normative perspective even if his sole aim is to complete his empirical picture. We have the normative perspective and

have to respect it as an empirical ideal even in doing empirical psychology. Hence, the normative first-person perspective not only cannot be eliminated *in toto*; this was already clear from the unavoidability of normative questions mentioned at the beginning of this paper. Rather, the subject's normative perspective cannot even be eliminated from the observer's empirical perspective. You cannot complete empirical psychology without engaging in normative considerations.⁹ Of course, this observation spreads from psychology to all human affairs, economics, social and political science, and so on. Therefore, even from the empirical perspective one is committed to the normative perspective. It would not be correct to distinguish one of the perspectives as the primary one.

When we emphatically claim that we have a free will and that our actions are first causes, we speak the truth from our normative perspective. And we are bound to acknowledge this when we change to the empirical perspective.

NOTES

1. This paper emerged from a commentary on Henrik Walter's *Neurophilosophy of Free Will* (1998) presented at the seventh Pittsburgh-Konstanz Colloquium in Konstanz in May 2005. Although I sense the wiggling strategy in this book, it is a brilliant essay based on current research in neurophysiology and cognitive science, explaining the extent to which compatibilism may be realized in an empirically informed way.

2. All earlier formulations of decision theory conformed to this principle. Richard C. Jeffrey ([1965] 1983, chap. 5) was the first to violate it. This is how my attention was drawn to it, and I still think this is the basic point where Jeffrey's decision theory goes wrong. Isaac Levi (1986, sec. 4.3, in particular pp. 65–66) concurs with the thesis that deliberation crowds out prediction. Since then, the thesis is under debate; see for example Rabinowicz 2002. Indeed, the point is subtle; self-prediction is not obviously absurd. My most recent view on the matter is presented in Spohn 2003, sec. 4.

3. I have developed such an account in many papers, most extensively perhaps in Spohn 1990 and 2006. My view on how to extend it to practical decision models (or from Bayesian nets to influence diagrams) is, however, only presented in Spohn 1978, secs. 3.2 and 3.3. Similar accounts may be found in Spirtes et al. 1993, chap. 3, in particular sec. 3.7.2; and Pearl 2000, chap. 4. In Spohn 2001 I briefly discuss the similarities.

4. I feel quite close to his position, also because I concur with his way of arguing

from a decision-theoretic background. However, I do not follow his argument on pp. 69–78, where he feels forced to accept that the difference between naturalism and humanism already has roots in ontology. See my second part here for my diverging view.

5. I should mention that there are also strands in the logic of actions emphasizing the uncausedness of actions; see Åqvist 1974 or von Kutschera 1986. Again, though, the relation to our present topic does not seem clearly developed.

6. For a detailed account of truncation see Pearl 2000, sec. 3.2. Truncation is similar to what Spirtes et al. (1993, sec. 3.7.2) describe as manipulation. In fact, the truncation procedure is already found in Spohn 1978, sec. 5.2. See Spohn 2001 for more detailed comparative remarks. I have discussed some interesting and consequential subtleties concerning truncation in Spohn 2003, sec. 4.3.

7. This is a point of some significance. There is a widespread view, unfortunately not only among those adverse to decision theory but also among decision theorists themselves, that decision theory may be taken as a computational theory, as about the subject's deliberative procedures. I believe this is a serious misunderstanding. Decision theory makes claims about the relation between beliefs, desires, and actions, but not about the computational relating. Similarly, logic tells not how we should proceed to reason, but how we may proceed to reason and whether or not the result of our reasoning is correct.

8. Our normative and empirical theorizing is thus characterized by a complicated twofold reflective equilibrium that I have discussed in some detail in Spohn 1993.

9. Of course, I do not want to suggest that the normative part of psychology is in any way dominant. For instance, if psychologists investigate the complex phenomenon of dyslexia, any kind of normative theorizing would be beside the point. The same holds for large parts of psychology. I only claim that one can never exhaust psychology in this spirit.

One must not forget that each empirical inquiry is full of normative issues. The inquiry has to conform to methodological norms that may well be debatable, and each inquiry starts with the decision to invest efforts in this rather than that issue. These are often difficult and far-reaching normative questions. It should be clear, however, that the normative entanglement of psychology I am arguing for goes far beyond this general normativity of empirical inquiry.

REFERENCES

- Åqvist, Lennart. 1974. A New Approach to the Logical Theory of Actions and Causality. In *Logical Theory and Semantic Analysis*, ed. S. Stenlund, 73–91. Dordrecht: Reidel.
- Dennett, Daniel. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, Mass.: MIT Press.
- Jeffrey, Richard C. [1965] 1983. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.

- Kant, Immanuel. [1785] 1906. *Grundlegung zur Metaphysik der Sitten*. Ed. Karl Vorländer. Hamburg: Meiner Verlag.
- Kutschera, Franz von. 1986. Bewirken. *Erkenntnis* 24:253–81.
- Levi, Isaac. 1986. *Hard Choices: Decision Making under Unresolved Conflict*. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Nida-Rümelin, Julian. 2005 *Über menschliche Freiheit*. Stuttgart: Reclam.
- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Rabinowicz, Wlodek. 2002. Does Deliberation Crowd Out Self-Prediction? *Erkenntnis* 57:91–122.
- Spirtes, Peter, Clark Glymour, and Richard Scheines. 1993. *Causation, Prediction, and Search*. New York: Springer.
- Spohn, Wolfgang. 1977. Where Luce and Krantz Do Really Generalize Savage's Decision Model. *Erkenntnis* 11:113–34.
- . 1978. *Grundlagen der Entscheidungstheorie*, Kronberg/Ts.: Scriptor. Pdf version available at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/Mitarbeiter/spohn_books.shtml.
- . 1990. Direct and Indirect Causes. *Topoi* 9:125–45.
- . 1993. Wie kann die Theorie der Rationalität normativ und empirisch zugleich sein? In *Ethik und Empirie: Zum Zusammenspiel von begrifflicher Analyse und erfahrungswissenschaftlicher Forschung in der Ethik*, ed. L. Eckensberger and U. Gähde, 151–96. Frankfurt am Main: Suhrkamp.
- . 2001. Bayesian Nets Are All There Is to Causal Dependence." In *Stochastic Dependence and Causality*, ed. M. C. Galavotti et al., 157–72. Stanford, Calif.: CSLI Publications.
- . 2003. Dependency Equilibria and the Causal Structure of Decision and Game Situations. *Homo Oeconomicus* 20:195–255.
- . 2006. Causation: An Alternative. *British Journal for the Philosophy of Science* 57: 93–119.
- Walter, Henrik. 1998. *Neurophilosophie der Willensfreiheit*. Paderborn: Mentis. English translation: *Neurophilosophy of Free Will*. Cambridge, Mass.: MIT Press, 2001.
- Wright, Georg Henrik, von. 1971. *Explanation and Understanding*. Ithaca: Cornell University Press.