

“Part of that force that always wills the evil  
and always produces the good”  
On a Devilish Incoherence

Peter Baumann

When Mephisto was asked by Faust, "Well now, who are you then?" ("Nun gut, wer bist Du denn?"), he gave the well-known answer, "Part of that force that always wills the evil and always produces the good" ("Ein Teil von jener Kraft, die stets das Böse will und stets das Gute schafft": Goethe, *Faust*, 1334-1336). This answer raises some important questions which in turn lead to interesting answers. Let us start by looking at the content of Mephisto's utterance.

1. What Mephisto Is Saying: The First Part

For simplicity's sake we may ignore the first few words of Mephisto's answer: "part of that power". Nothing substantial changes if we assume that Mephisto ascribes to himself, too, what he ascribes to "that power" (if not, then we would just have to talk about "that power" instead of Mephisto).<sup>1</sup> We can assume that Mephisto also wants to make the following assertion: "I always will the evil and always produce the good".

The expression "will" does not refer to wishes or even mere wishes (compare the German "wollen" to "wünschen"). To will something (e.g., evil) involves being moved towards action in a relevant way (e.g., to do evil). Apart from this, Mephisto's utterance states or at least suggests a clear connection between the attitude of willing and the production of relevant consequences. Should we go one step further and interpret "will" as "intend"? The wording of Mephisto's utterance speaks against this move. It is important to keep in mind that intending differs from willing (see Bratman 1987). Even though someone who intends to bring it about that  $p$  also wills that  $p$ , the reverse does not hold: Someone can will that  $p$  (e.g., that he gets a tan) without intending to bring it about that  $p$ . We should thus stick with the following paraphrase of Mephisto's remark:

“I always will the evil and always bring about the good”

rather than the following one

“I always intend to do evil and always bring about the good”.

If we went with the latter instead, then there would be a quick way of showing Mephisto to be incoherent. An intention to bring about X requires, amongst other things (like the belief that one has a good chance of bringing about X), the absence of the belief that one will fail to bring about X. However, the second conjunct above (“I always bring about the good”) expresses just such a belief. The second part of the conjunction expresses a belief which falsifies the first part of the conjunction (interpreted as being about intentions). To be sure, it can be true of an agent that he intends to bring about X but always fail to do so. However, he cannot coherently believe this. There is a certain resemblance with Moore’s Paradox (see Moore 1959, 175-176): “It’s raining but I don’t believe it” may be true but cannot be believed coherently. Let us then rather stay with the first paraphrase above: “I always will the evil and always bring about the good”.

The expression “evil” (or “to do evil”) refers to the object or content of the will. Mephisto is thus not saying merely that his will is directed towards something which turns out to be evil but rather that he wills the evil as evil. Related to this is another important point. Mephisto’s will is evil because he wills that bad things happen; it is not evil, say, in a Kantian sense: because the will’s maxim has a certain form (see, e.g., Kant, *Grundlegung zur Metaphysik der Sitten*). The wording of Mephisto’s assertion speaks against such a Kantian interpretation. Hence, we can paraphrase Mephisto’s utterance in the following way:

“I always will that bad things happen and always bring about the good”.<sup>2</sup>

Sentences about one’s will like the above can be expressions of one’s will or truth-apt (self-) ascriptions of a will (see, e.g., for the parallel case of intentions: Anscombe 1958). Here we are dealing with the latter case. The location and scope of “always” indicates this. “I will that always *p*” can be taken as an expression of a will, too, but “I always will that *p*” can only be taken as a description of a will. Mephisto is self-ascribing the systematic and continuous will for bad things; he is not self-ascribing or expressing the will that bad things happens systematically and continuously. On another occasion he might be tempted to do that, too, but not on this one, when he is asked who he is.

Mephisto attributes to himself what one can call an “evil will”. We do not need to discuss the question here whether there can be an evil will for human or non-human actors. Let us suppose this is possible. We can also put aside another potential worry. Mephisto is not saying that he wills what others think is bad but rather that he wills what he himself thinks of as bad. But in what sense can one will something that one thinks is not good but bad? Doesn’t – as Plato’s Socrates argued a long time ago (see Plato, *Meno*, 77b-78b) - willing that *p* imply that the agent holds that *p* is good in some important sense (and in his own lights)? And equivalently: Doesn’t believing that *p* is

bad imply that the agent doesn't will (in some important sense) that  $p$ ? We can leave this problem, too, aside, because our main focus will not be "good" and "evil" or "bad" specifically but on contraries more generally.

## 2. What Mephisto Is Saying: Adding the Conjunct

Apart from the self-ascription of an evil will Mephisto also asserts that he always brings about the good. He is pointing out that the actions he performs with evil will bring about unwelcome good consequences. He does not mean only to say that there are also some good consequences within the total set of the consequences of his actions: some good consequences apart from the not so good, bad or neutral consequences. Rather, we should interpret him as either characterizing the consequences of each of his actions as predominantly or overall good or, alternatively, as characterizing the consequences of each of his actions as completely good. I do find the second interpretation more plausible. The expression "always produces the good" ("stets das Gute schafft") has to be contrasted with "always produces some good" ("stets etwas Gutes schafft"); the definite article at least suggests exclusivity of good outcomes. Apart from that, the second interpretation also fits better with Mephisto's view of his own nature as one who always negates ("der Geist, der stets verneint": Goethe, *Faust* 1338). For Mephisto, there are no "mixed" consequences (some good, some bad, some neutral). Similarly, we shall assume that Mephisto is ascribing only a will to himself that is exclusively evil and without any component of a good or neutral will (this excludes a divided or "mixed" mind, as in cases of weakness of the will where the agent desires the bad while also desiring the good in some sense).

It is not that unusual that in a particular case an actor's will not only does not succeed but also accomplishes something contrary to his will. That this is possible is a fundamental aspect of an agent's predicament. One can call this discrepancy between a will and a contrary consequence of the corresponding action a "perverse condition". Mephisto's claim is puzzling and provocative because of the claimed exceptionlessness of the perverse condition: "I always will that bad things happen and always bring about the good". Strictly (and pedantically) speaking Mephisto leaves it open whether there is a causal link or not (and then only a mere correlation) between his will and the outcomes of his actions. However, it is hard to see how there could not be such a causal link. Apart from that: The good, he claims, is produced by himself, that is, by his actions, which in turn are always guided by his evil will. Hence, we may assume that his evil will leads to actions which bring about or cause good outcomes.

## 3. Mephisto's Incoherence and the Impossibility of Extreme Perversion

At this point one should really wonder whether it is even possible that an agent suffers from an exceptionless perverse condition. How could Mephisto possibly be right? Let us, again, look at the two conjuncts in Mephisto's claim "I always will that bad things

happen and always bring about the good” in separation. He makes two (sub-) claims. First:

(1) I always will that bad things happen as a consequence of my actions.

(Mephisto does not just wish that bad things happen in some way or other; he wills it in an “active way” (see above), desiring it to happen as a consequence of his own actions). Second:

(2) All my actions have good consequences.

(2) plus the exclusion of mixed consequences (see above) and the trivial truth that what is good cannot be bad entail

(3) None of my actions has bad consequences.

Now, there is no reason to assume that only when Mephisto is not acting does he believe or think that he always wills that bad things happen and always brings about the good. Rather, we may assume that he always acts in the light of this belief. So, whenever he acts he expects only good consequences. This is important because the following principle is very plausible:

(P) If an agent performs some action willingly and expects it to have only a certain kind of consequences (good or non-bad ones), then the agent wills to produce this kind of consequences.

To avoid misunderstandings: (P) does not claim that there are no unwanted or unintended consequences of actions; it only makes a claim about expected kinds of consequences.

Now, given that Mephisto has a will and performs willful actions, it follows from (P) together with the fact that Mephisto always believes (3) that (upon brief reflection) he also believes:

(4) I always will that nothing bad happens as a consequence of my actions.

Thus, Mephisto is committed to the conjunction of (1) and (4):

(5) I always will that bad things happen as a consequence of my actions, and I always will that nothing bad happens as a consequence of my actions.

This kind of conjunction is true of any agent who has a belief like Mephisto.

This conjunction is not yet a contradiction (if we don't add "only" to "always") but even if we suppose that Mephisto is not inconsistent he already shows a practical deficit. He has an ambivalent will: He is undecided and wavers between mutually incompatible motives or wills, his good will and his evil will. This is something we should not have expected, especially not from a devil. This should rather be Faust's problem: "Two souls, alas, are dwelling in my breast" ("Zwei Seelen wohnen, ach! in meiner Brust": Goethe, *Faust*, 1112).

There is a contradiction "in the vicinity", however: The ambivalence of Mephisto's will follows from a couple of assumptions including one to the effect that Mephisto has an "un-mixed" will (see above). Given the latter assumption, we can sharpen (5) by adding "only" to "always":

(6) I always only will that bad things happen as a consequence of my actions, and I always only will that nothing bad happens as a consequence of my actions.

Given, again, that Mephisto really has a will and performs willful actions it follows from the second conjunct of (6) that (cancelling any Gricean implicatures)

(7) I sometimes only will that nothing bad happens as a consequence of my actions.

Given the incompatibility of only willing the bad with only willing the good, the first conjunct of (6) implies that

(8) I never only will that nothing bad happens as a consequence of my actions.

(7) and (8) together give us a contradiction! Mephisto is simply inconsistent: not openly but what he says entails (given reasonable interpretation) a contradiction.

One wonders why Mephisto would come up with an inconsistency like that, and also why he presents himself as possessing a deeply ambivalent will. Is he being ironic? Does he want to tease, test or tempt Faust (or the rest of us)? Is he playing the devil's

advocate (but how that?)? And if he is serious: How could he continue acting? Or does he just want to get away with a pseudo-answer and not identify himself (which would fit with his reluctance to give his name a moment earlier: see Goethe, *Faust*, 1327-1330)?

#### 4. Generalization: Why We Have to Be Successful at least Sometimes

Whatever we might think resolves this puzzle, it is worth adding here that similar contradictions as well as ambivalences of the will can be detected in cases where the claimed consequences of one's actions are not contraries of what was willed but where the will is just not realized or not fully realized. In such cases Mephisto could say something like "I always will the evil and always bring about what is not evil" or "I always will the evil and always bring about something that falls short of (complete) evil". Perverse conditions are also not restricted to a will directed at good and evil (or bad) but can arise for wills of any content. And there can also be an inverse Mephisto who claims that he always wills the good and always brings about bad consequences. Finally, the perversity of Mephisto's condition does not depend on him always willing the same kind of thing; rather, whatever he wills, he is always in for disappointment. Hence, we can generalize the idea of an exceptionless perverse condition in the three indicated ways. Mephisto could then also assert that

"Whatever I will to happen, the consequences of my action never match my will".

An agent suffers from a general and exceptionless perverse condition just in case he never manages to bring about what he wills to bring about.

Both Mephisto's original and the generalized assertion cannot be true because they each entail a contradiction, given that Mephisto believes that he suffers from the perverse condition (he doesn't even need to know this: mere belief is sufficient). No agent can be in an exceptionless perverse condition if he believes he is in such a condition. Even if the agent has no beliefs about the content of his will (and thus not about the perversity of his condition) and only has beliefs about those aspects of the consequences of his actions in general that are relevant to his will, the agent still cannot be in an exceptionless perverse condition (see the support for (4) above and what follows). In other words, given such beliefs about the consequences of one's actions in general, one cannot be in an exceptionless perverse condition: At least some of one's many particular "wills" have to be successful. One can only be in such a condition if one lacks beliefs about the consequences of one's actions in general. A corollary of all this says that given certain beliefs about the consequences of one's

actions in general, one can infer “in the armchair” that some of these consequences meet one’s will.

### 5. Going even further: Panglossian Temptations?

Can we go even further? What if Mephisto said that “Whatever I will to happen, the consequences of my action almost never (but not never) match my will”? Can one be in such a “less extreme” perverse condition? Perhaps one can, even given beliefs about the consequences of one’s actions in general. Perhaps then it is also possible that one finds oneself in a condition where one’s will only often (but not never or even almost never) remains unfulfilled. It certainly is possible that one sometimes faces a perverse condition. However, it also seems plausible to assume that the more often one faces perverse conditions, the more one’s agency erodes: the more one loses confidence and the energy to act. This would suggest that insofar as one acts consciously at all one must have a great deal of success as far as the willed consequences of one’s actions are concerned. I will end with a few remarks sketching an argumentative strategy for such a moderately “panglossian” conclusion.

Consider the possibility that whatever an agent wills to happen as a consequence of his actions, more often than not the consequences of his actions do not match his will. (I am leaving aside the thorny problem of how to individuate and count wills or acts). Suppose further that the agent has the corresponding belief about the consequences of his actions. If the agent has no way to tell cases in which he does get his will from cases in which he doesn’t get his will, the agent might be entitled to use a principle of indifference and, given that he is minimally rational, assign a probability  $>.5$  to the possibility of failure of his original will in each individual case. But since the agent performs willful actions in each individual case, we then also have to ascribe to the agent in each such case a will to bring about such failure rather than success of his original will. The same kind of ambivalence and even inconsistency as in Mephisto’s exceptionless perverse condition follows in this case, too (we don’t need to go through the details).

If, on the other hand, the agent has a way to tell cases in which he does get his will from cases in which he doesn’t get his will, then he will, as a minimally rational agent, continue to act willfully in the cases where he thinks he has a good chance of success and give up acting at all in the cases where he thinks he has no good chance of success. If his ideas about probable success are somewhat realistic, then he will only fail to get his will in a minority of cases. If his ideas are unrealistic and often false, then he will often fail to get his will. It is, however, not easy to imagine how he could not find out rather quickly that he is wrong; after correcting his ideas, he would be back with the former, better condition.

So, the upshot of these final remarks is that, given certain beliefs of the agent, it cannot happen that more often than not agents don’t get their will.

Davidson once argued that most of our beliefs must be true (and “successful” if truth is the aim of belief) (see Davidson 1984, 199). We can’t go quite as far with respect to willful action. So, Armstrong might go a bit too far when he remarks that “unless objectives were regularly achieved, there would be no place for the notion of objective” (Armstrong 1968, 144). However, something a bit weaker would seem correct. One final caveat is important to keep in mind. Even if our claims about necessary success go further and beyond Mephisto’s exceptionless perverse condition, and even if we can claim that agents cannot, given certain conditions, face more failure than success, this still does not mean or entail that we get our will in the most important cases. In this sense, too, we would have only reached a moderate panglossian conclusion.<sup>3</sup>

Peter Baumann  
 Swarthmore College  
 pbauman1@swarthmore.edu

#### NOTES

<sup>1</sup> One might object that the latter does not follow from the fact that Mephisto is part of a force that always wills the evil and always produces the good. The part could differ significantly from the whole in this respect. It might thus still be false that Mephisto himself, too, always wills the evil and always produces the good. However, one should reply: Could it, really? First, it is very hard to see how Mephisto could, for instance, never will the evil (perhaps even always will the good) and always produce the good, or how he could, in contrast, always will the evil and always produce it. How could he then still be “part of that force” from which he would be so radically different? He would then rather be a part of a very different kind of force: a divine force in the former case or a more traditional devilish force in the latter case. Second, if the difference between Mephisto and “that force” is only small – like in the case where Mephisto himself always wills the evil but only almost always produces the good (also see section 5 below on such cases) –, then it is hard to see why we should ascribe such a small divergence from “that force” to Mephisto. Apart from all this, the text of Goethe’s *Faust* strongly suggests an identification of Mephisto and “that force” in the relevant respects. We can be confident that Mephisto would have mentioned any relevant difference from the force. Finally, even if Mephisto should have fallen from “that force”, too (imagine!) we could then just switch from talk about Mephisto to talk about that force: It is an agent, too, and what matters here are problems of agency in general.



<sup>2</sup> Are we changing the topic here from “evil” to “bad”? I don’t think so. The term “evil” serves as a characterization of the will, the term “bad” as a characterization of the content or object of such a will.

<sup>3</sup> For comments and discussion I am grateful to my parents as well as to Reinhard Baule, an audience in Leipzig (Germany), and anonymous referees.

#### REFERENCES

Anscombe, G.E.M. 1958, *Intention*, Oxford: Blackwell.

Armstrong, David Malet 1968, *A Materialist Theory of the Mind*, London & New York: Routledge & Humanities Press.

Bratman, Michael E. 1987, *Intention, Plans, and Practical Reason*, Cambridge/MA: Harvard University Press.

Davidson, Donald 1984, “The Method of Truth in Metaphysics”, in: Donald Davidson, *Inquiries into Truth and Interpretation*, Oxford: Clarendon, 199-214.

Goethe, Johann Wolfgang von, *Faust. Eine Tragödie*, in: Goethes Werke. Hamburger Ausgabe (ed.: Erich Trunz), München: C.H. Beck 1982-2008, vol.III.

Kant, Immanuel, *Grundlegung zur Metaphysik der Sitten*, in: Immanuel Kant, *Gesammelte Schriften*, Berlin: G. Reimer 1902ff., vol.IV, 385-463.

Moore, George Edward 1959, "Russell's Theory of Descriptions", in: George Edward Moore, *Philosophical Papers*, London/ New York: Allen & Unwin/ Macmillan, 151-195.

Plato, *Meno*, in: *Five Dialogues* (tr.: G.M.A. Grube & John M. Cooper), Indianapolis: Hackett 2002.