

FINKING FRANKFURT

Daniel Cohen & Toby Handfield†*

1. Laertes is about to stab Hamlet. Claudius, watching from the wings, desperately wants Laertes to stab Hamlet and has implanted a device in Laertes' brain that, if activated, will force him to stab Hamlet. Claudius intends to activate the device if it appears that Laertes will change his mind at the last moment. Laertes, however, stabs Hamlet of his own accord and Claudius's device is never activated.

By appealing to examples such as this, Harry Frankfurt (1969) presents a powerful objection to the view that moral responsibility requires the ability to do otherwise (*the principle of alternative possibilities*, or PAP). Frankfurt's argument relies on two claims: firstly, that Laertes is responsible for stabbing Hamlet and secondly, that Laertes lacks the ability not to do so. The former claim seems obviously true. The latter claim is, however, more controversial.

In particular, the latter claim depends on how 'ability' is understood. In this paper, we shall discuss an analysis of ability in terms of agents' dispositions. Relying on a dispositional analysis of ability, Michael Smith (1997, 2003) has argued that agents in Frankfurt examples, such as Laertes, should be thought of as possessing the ability to do otherwise, contra Frankfurt. We will grant that Smith is right about agents such as Laertes, but argue that another classic Frankfurt case evades Smith's analysis. Frankfurt's *willing addict* (1971), in particular, cannot be shown, on Smith's account, to possess the ability to do otherwise. Provided one shares the intuition that the willing addict is indeed responsible, our argument suggests that Smith fails, ultimately, in his defence of PAP.

We do not intend to argue here that *no* account is available according to which the willing addict possesses the ability to do otherwise. We will only argue that no dispositional account is available. Nevertheless, if we are right, the burden of proof will fall on the defender of PAP to provide an account that does not, however tacitly, rely on dispositions.

2. Smith (2003) suggests that a variety of puzzle cases, including Frankfurt's examples, can be explained in terms of a *dispositional* analysis of the ability to do otherwise. So, to the extent that Laertes may be said to have the disposition not to

* *Philosophy Program, RSSH, Australian National University and School of Philosophy & Bioethics, Monash University, email: cohen at coombs.anu.edu.au.*

† *School of Philosophy & Bioethics, Monash University, email: toby.handfield at arts.monash.edu.au.*

stab Hamlet – despite the presence of Claudius and his device – we may salvage the thought both that Laertes has the *ability* to refrain from stabbing and that this ability is necessary for him to be responsible for his behaviour.

Simply, Smith’s view is that Laertes is responsible only if he possesses two crucial capacities, one being the capacity to form correct evaluative beliefs (a capacity for moral judgment), and the other being a capacity to bring one’s actions into accord with one’s evaluative beliefs (a capacity for moral action, one might say.)

These capacities are akin to dispositions. To ascribe a capacity to someone is to say something about what they could do, without necessarily saying what they will do. It is to locate agents in modal space: to say something about how things will go under various possible circumstances. If, as we may imagine the case, Laertes’ behaviour does not accord with his evaluative beliefs, Laertes’ possession of the capacity for moral action means that it was in his power to act otherwise: to act in accordance with his evaluative beliefs and to refrain from stabbing Hamlet.

Dispositions are typically thought of as dispositions *to* give a certain manifestation, *if* a certain stimulus phenomenon occurs. Smith’s capacities have clearly described manifestation conditions – forming correct evaluative beliefs, acting on one’s evaluative beliefs – but he is chary of giving an account of the stimulus conditions for these capacities. This is not to say that Smith’s capacities need be brute propensities, which can manifest without any stimulus condition at all. Rather, the idea appears to be that the capacities do not have any very well understood, and perhaps only indeterminate, stimulus conditions. Smith does say, however, that the capacity must be suitably “multi-track” (2003, p. 123). That is to say: they must manifest similarly under a reasonably wide range of similar circumstances.

To try and characterize in counterfactual terms, then, what is meant by saying that Laertes has the capacity to do otherwise, one might say: *Had things been otherwise, in a wide variety of suitably minor ways, Laertes would have acted otherwise.* Of course, in the situation described, this counterfactual is probably false, because of the presence of Claudius. Thus, if one relies upon a counterfactual characterization of Laertes’ capacities, it is somewhat dubious to claim that Laertes has the capacity to do otherwise in this case.

However, it has been known for some time that dispositional properties cannot be analysed in terms of simple counterfactuals, of the sort given above. Smith draws on this recent work on dispositions to suggest that certain counterexamples to the conditional analysis of dispositions – finks and masks – are of precisely analogous structure to Frankfurt’s cases.¹ Finks and masks show that the simple conditional analysis of dispositions is false, but they don’t show that objects lack dispositions, even when finks and masks are present. Analogously then, Smith suggests that

¹ For another, undeveloped, suggestion along these lines see Vihvelin 2000, note 34.

Frankfurt cases show that simple conditional analyses of the capacities for moral judgment and moral action are false, but that the cases do not show that we lack these capacities.

A fink is a feature of a situation such that, were a disposition to be triggered by its stimulus, the fink would very quickly act to remove the disposition before it could be manifested (Martin 1994). David Lewis's influential example of a fink is a sorcerer who wishes to protect a glass from breaking (1997). If the glass were struck, the sorcerer would very quickly change the molecular structure of the glass so that it would no longer be fragile. Thus it would not be true to say of the glass that, if struck, it would break. Nonetheless, insofar as the glass is never struck, there is a strong intuitive appeal in the thought that the glass is fragile.

A mask is like a fink in that it acts to prevent the manifestation of a disposition, but rather than removing the disposition itself, a mask (or antidote) interferes with the manifestation process (Johnston 1992, Bird 1998). A poison is disposed to cause death if ingested. For some poisons, however, if I ingest them I shall not die, because I will take an antidote which prevents the poison from manifesting its disposition.

By considering Laertes in light of these possibilities, we may now understand how Smith defends the dispositional analysis of ability. Laertes is like the fragile glass protected by the sorcerer. Just as the glass remains fragile, at least when it is not in fact dropped, Laertes may be thought to possess the ability to do otherwise, at least while Claudius's device is not activated. Claudius, like the sorcerer, is a fink because he will destroy Laertes' disposition not to stab Hamlet, but only in those worlds where his disposition would otherwise have been manifested.²

Even though both the glass and Laertes will not manifest their dispositions in the presence of a fink, they may still be said to have their dispositions, when not actually finked. This is why, in identifying Laertes' ability, Smith says that we must 'abstract away' from all features external to the relevant features of his brain, at least when these features don't interfere with his behaviour (Smith 2003, pp. 126–7). The irrelevant features will, of course, include Claudius and his device.

We welcome Smith's account of the phenomena. However, there is crucial work to be done in exploring the range of cases which can be analysed in this way. The appeal to examples only shows that some Frankfurt situations are compatible with the possession of the disposition to do otherwise. It does not show that *all* Frankfurt situations are so compatible. Is it possible that some Frankfurt situations are "radical", such that their obtaining suffices to *eradicate* the disposition to do otherwise?

² Depending on whether Claudius's device *removes* Laertes' disposition to do otherwise, or whether the device simply prevents this disposition from being manifested, Claudius may in fact be either a fink or a mask. We shall assume he is a fink, but nothing of importance turns on things being one way or the other.

We here argue that there could indeed be such radical finks. Moreover, we show that the possibility of such finks throws up a problem for Smith's dispositional defence of PAP. The willing addict, as we shall argue, is intuitively responsible for her behaviour, yet she cannot be said to possess the ability to do otherwise, at least on Smith's dispositional analysis.

3. In philosophical discussion, finks are typically conceived of as something extrinsic to the bearer of the disposition. The presence of Lewis's sorcerer, for instance, is not an intrinsic feature of the glass. Similarly with masks: it is not an intrinsic property of a given poisonous substance to be such that, upon being ingested, the poisoned party will also ingest an antidote. Clearly this is an extrinsic property of the substance, because intrinsically perfectly similar substances could lack this property.

Could there be *intrinsic* finks for a disposition? Following Sungho Choi (2005), we suggest not. Consider a glass made of a substance which is intrinsically such that, when struck, it acquires a very durable molecular structure. Consequently, if it were struck the glass would not break. One might have thought that this is just another fink. The glass is fragile, but it is in a situation such that, if exposed to the stimulus, it would cease to be fragile.³

We believe this would be a mistake. It is intuitively plausible that such a glass is better described as *non-fragile*. If one has no clear intuition about this glass, consider Choi's heuristic (pp. 499–500) for determining whether an object has an intrinsic disposition: does any intrinsic duplicate of the object, subject to the same laws of nature, obviously possess the disposition? The object may have intrinsic duplicates of which it is true to say that, if struck, they would break. But all such possible glasses will have this property in virtue of some extrinsic factor, such as a fink or mask. None of these objects *obviously* have the disposition, given the complication of these external factors. Given this, we suggest it is very plausible to deny that the glass is fragile at all.

This pattern appears to generalize. If an object is alleged to have a disposition, but also to have an *intrinsic* fink or mask to that disposition, it is very difficult to see why we should accept this description, rather than simply deny that the object has the disposition at all.

We make the following conjecture, then: *dispositions cannot be intrinsically finked or masked*. If an object with a disposition were to acquire an intrinsic property which finked or masked the disposition, then it would simply cease to have the disposition. Or, to put it another way, intrinsic finks and masks are radical: upon instantiation they eradicate the original disposition.

³ This appears to be what Lewis (1997, p. 157) thought. He suggests a glass could be both fragile and non-fragile, provided one of those dispositions is finkish, and the other is not.

4. The case of Laertes and Claudius is a clear example of an extrinsic mask. An intrinsic duplicate of Laertes, subject to the same laws of nature, may clearly possess the disposition not to stab Hamlet. Therefore there is no difficulty in maintaining, with Smith, that Laertes possesses the disposition not to stab Hamlet, and in virtue of this, the ability not to do so.

Frankfurt's willing addict (1971), however, is less easily assimilated to this paradigm. The willing addict has a disposition to crave narcotics. This craving, moreover, is strong enough to *force* the addict to ingest narcotics. (It is in virtue of this that she is, indeed, addicted). This disposition will be triggered whenever she goes without narcotics for a significant period of time. The addict, however, never goes without narcotics for that period of time, because she willingly chooses to ingest them at regular intervals. We may imagine that, due to her constant ingestion of narcotics, the willing addict may not even realise that she is addicted. Frankfurt claims that, like Laertes, the willing addict is responsible for her behaviour, despite not being able to do otherwise. We have seen that Smith is able, however, to show that Laertes may, in fact, possess the ability to do otherwise, if abilities are understood dispositionally. We will argue that Smith cannot make this claim in the case of the willing addict. On Smith's account, the addict will be able to do otherwise only to the extent that she possesses the disposition not to ingest narcotics, a disposition that might be finked (or masked) by her addiction (Smith 1997). However, does the willing addict possess such a disposition? This depends crucially upon the relation between the addict and her addiction.

Either the addiction is intrinsic to the addict or it is extrinsic to the addict. Neither option is trouble-free.

The first option will seem clearly to follow if we think of the addict as at least partly constituted by her body. An intrinsic duplicate of the addict's body, subject to the same laws of nature, will plausibly share the disposition to ingest narcotics. If this is the case, however, Smith cannot say that the willing addict has the ability to do otherwise: whatever disposition the addict may have possessed to refrain from taking drugs has been eradicated by the radical mask of her addiction. Assuming that the addict is indeed responsible for her behaviour, Smith cannot thus claim that the ability to do otherwise is necessary for moral responsibility. We seem forced to deny PAP, vindicating the original purpose of Frankfurt's examples.

Smith might, however, bite the bullet. He may say that Laertes is responsible because he does possess the ability to do otherwise: his disposition to do otherwise is finkish, but is nonetheless present. The willing addict's ability, on the other hand, is not simply finkish, but actually eradicated by her addiction. On this basis, the addict cannot be held responsible for her behaviour. This move is problematic because the intuition that the willing addict is responsible appears fundamentally *on par* with the intuition that agents such as Laertes are responsible. (Plausibly, they are instances of the *same* intuition, so that they stand or fall together.) Denying that the willing addict is responsible is only plausible to the extent that one also denies that Laertes

is responsible. Any putative analysis of responsibility must, moreover, respect this parity of intuitions.⁴

Perhaps, however, Smith may defend the intuition that the willing addict is responsible, despite lacking the ability to do otherwise *at the time of her addiction*. That is, he may argue that the willing addict is responsible for her addictive behaviour because she freely chose to become addicted at some time before she was addicted. On this view, Laertes and the willing addict are only superficially on par, for the willing addict is not strictly responsible for her addictive behaviour as such, but rather only for some earlier behaviour that caused her currently to behave as she does.⁵

This suggestion is problematic for a couple of reasons. Firstly it is not obvious that a free choice to initiate some chain of events will always render one responsible for all the events that follow, unless it were reasonably foreseeable that those events would follow. However, our intuitions about the responsibility of the willing addict for her behaviour don't seem contingent on any assumptions about what she foresaw at the time she became addicted. Secondly, it's not clear that our intuitions about the responsibility of the willing addict *at all* depend on there being some time when she could have avoided being addicted. Even if the willing addict were *born* addicted to narcotics, as long as she always consumes them on her own volition (without, let's imagine, even realising she is addicted), it seems clear that she is responsible for her behaviour. It is clear, however, that this is not an intuition about her choice to become addicted. Rather, it appears that the willing addict's deliberative choice to consume narcotics, by itself, is sufficient for her to be responsible.

The second option – saying that the willing addict's addiction is not an intrinsic part of her – is perhaps more superficially attractive, then. But on what basis would we say that the addiction is not intrinsic to the addict? Any psychological duplicate of the addict would surely also share the same addiction. Perhaps psychological duplication is too coarse and inclusive a relation to capture only the intrinsic features of a moral agent, as opposed to the intrinsic features of the behavioural system in which moral agency is realised. Perhaps by taking psychological duplicates, we are including features of the merely “phenomenal self”, in addition to those of the “noumenal self”. Smith himself suggests that it is just certain “relevant properties” of the brain that are to be duplicated in identifying an agent's capacities (Smith 2003, p. 122). However, it is not clear what reasons could be given for duplicating only the non-addicted parts of the willing addict's brain. The suggestion that only the non-

⁴ Smith may, of course, deny that Laertes is responsible, but this would be substantially to abandon the project of developing a compatibilist analysis of ability. If Claudius's counterfactual intentions are analogous to causal determinism, they would appear equally to preclude the ability to do otherwise. But if one maintains PAP, as Smith does, then determinism would appear to preclude responsibility.

⁵ Thanks to an anonymous referee who suggested this argument.

addicted parts of the addict's brain constitute her "real self" (see Wolf 1990) would seem conveniently to beg the question. Even if the distinction between real and superficial selves could be made sense of, it's not clear why an agent's addictive desires won't be part of her real self.

Another way of pressing this worry is to compare the willing addict with the *unwilling* addict, who "hates his addiction and always struggles desperately, although to no avail, against its thrust" (Frankfurt 1971, p. 12). On the dispositional analysis of ability, if the willing addict is said to have the ability to resist temptation, then, by the same token, the unwilling addict must have this ability also. The unwilling addict, however, seems blameless for giving in to temptation. The most natural explanation for this is that the unwilling addict *lacks the ability* to resist temptation, but Smith cannot say this, insofar as he credits the willing addict with the same ability.⁶

5. In conclusion, the possibility of radical finks leaves Smith with an ugly dilemma. Either he can fall in line with Frankfurt and abandon the claim that the ability to do otherwise is a necessary condition for moral responsibility (PAP), or he can deny some of the key intuitive phenomena regarding the responsibility of the willing and unwilling addicts. We have not, however, argued against PAP. Rather, we hope to have shown that the ability to do otherwise, if it is thought to support judgements of responsibility – cannot be analysed dispositionally.⁷

⁶ There is some suggestion that Smith would claim that the unwilling addict does, in fact, possess the ability to resist temptation but that he deserves to be excused because the failure to exercise this ability does not explain his behaviour (Smith 2003, p. 127). This approach is problematic because it suggests that the unwilling addict, due to his addiction, is *unable* to exercise his abilities. If this is the case, however, surely we must say the same about the willing addict. And if the willing addict is not able to exercise her abilities, then she too deserves to be excused.

⁷ Thanks to Neil Levy, an anonymous referee, and an audience at the University of Bristol for helpful comments on an earlier version of this paper. Thanks also to Neil for our title.

References

- Bird, Alexander. 1998. "Dispositions and Antidotes". *The Philosophical Quarterly* 48: 227–34.
- Choi, Sungho. 2005. "Do Categorical Ascriptions Entail Counterfactual Conditionals?" *The Philosophical Quarterly* 55: 495–503
- Cullity, Garrett and Gaut, Berys (eds.). 1997. *Ethics and Practical Reason*. Clarendon: Oxford University Press.
- Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility". *The Journal of Philosophy* 66: 829–39.
- Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy* 68: 5–20.
- Johnston, Mark. 1992. "How to Speak of the Colours". *Philosophical Studies* 68: 221–63.
- Lewis, David. 1997. "Finkish Dispositions". *The Philosophical Quarterly* 47: 143–158.
- Martin, C. B. 1994. "Dispositions and Conditionals". *The Philosophical Quarterly* 44: 1–8.
- Smith, Michael. 1997. "A Theory of Freedom and Responsibility". In Cullity and Gaut 1997. Reprinted in Smith 2004.
- Smith, Michael. 2003. "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion". In Stroud and Tappolet 2003. Reprinted in Smith 2004.
- Smith, Michael. 2004. *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*. Cambridge: Cambridge University Press.
- Stroud, Sarah, and Tappolet, Christine (eds.). 2003. *Weakness of Will and Practical Irrationality*. Clarendon: Oxford University Press.
- Vihvelin, Kadri. 2000. "Freedom, Foreknowledge, and the Principle of Alternate Possibilities". *Canadian Journal of Philosophy* 30: 1–24.
- Wolf, Susan. 1990. *Freedom Within Reason*. New York: Oxford University Press.