

Explanationism about Freedom and Orthonomy*

This is the penultimate draft of the article forthcoming in The Journal of Philosophy. Please cite the official JoP version

Freedom and Orthonomy

It is part and parcel of our self-conception as agents that some of our choices and actions in life are free. This is the sense of freedom at play in our subsequent practices of attributing and accepting praise and blame and of imbuing with meaning our projects in life. It is the sense, for short, that tracks our impression of being responsible agents, agents whose actions can be attributed to them. If we are indeed free, moreover, then this is not, we usually assume, a brute fact with no further explanation. There is something about us or our relation to the world that *makes* us free. A crucial task for any metaphysics of freedom is therefore to inquire into the metaphysical basis of freedom. It is to ask in *virtue of what* we are free, if we are free, or what *grounds* our freedom, if there is anything to be grounded.

One famous type of answer to this question is that our freedom is grounded in the internal organisation of our agency (see for example Frankfurt 1971). However, an entirely different type of answer has recently come into fashion. According to this idea, what grounds free agency isn't the internal structure of our mind, but our mind's relationship to the external world. Importantly, "the world" on this view not only contains descriptive facts, but also a variety of normative demands bearing on our conduct, so-called normative reasons for action. Free agency is accordingly seen as the result of acting in possession of the ability to be "sensitive" to, or "respond" to, these demands. This ability has been called *orthonomy*, and I will refer to the relationship between the agent's action and the demands of the normative situation as *orthonomous relationship* (Wolf 1990, Pettit and Smith 1990, Pettit and Smith 1996, Smith 2003). Expressed in terms of the grounds of freedom, the basic idea of orthonomy accounts is this:

(G) Freedom facts are grounded in orthonomy facts.

This paper is about the nature of facts about orthonomy. It is about how we should understand the orthonomous relationship, if we want it to ground facts about free agency. For upon reflection, the idea that agents are sensitive or responsive to the demands of their normative situations is theoretically mysterious. What exactly does it mean to be responsive to demands?

* I am grateful to colloquium for theoretical philosophy FU Berlin and the colloquium at the University of Erlangen for invaluable feedback on precursor versions of this paper. I also owe thanks to a very helpful anonymous reviewer of this journal, who helped improve the paper considerably. Special thanks to Myriam Stihl. I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the project "Capacities and the Good", project Number: 439616221.

How exactly must an agent's action relate to their reasons in order for the action to be an instance of the relevant sensitivity? All we can say without further philosophical reflection is that we have in mind a kind of "robust rational acuity" with respect to reasons, the capacity to truly recognise and act *based* on them. But what does this acuity consist in?

I will argue for an *explanationist* answer to these questions, according to which the orthonomous relationship consists in the special explanatory relation between agent, reason, and action that we express when we give reasons-explanations of action. Thus, I will argue for:

(G-E) Freedom facts are grounded in explanatory facts.

The best way to see the merits of (E), however, is to engage in some philosophical story-telling about two extant candidates for the nature of orthonomy. The first is a causalist understanding:

(G-C) Freedom facts are grounded in causal facts.

Causalists hold that orthonomy facts are just causal facts. The orthonomous relationship between reason and action, they believe, comes down to a causal relationship (of the right kind). But (G-C) is false. It entails that there cannot be a contrast with respect to freedom and orthonomy without a causal contrast. But there can be (section 3).

The second (and perhaps most famous) extant candidate is a modalist understanding:

(G-M) Freedom facts are grounded in modal facts.

Modalists hold that the orthonomous relationship consists (or is marked by) modal companionship between reason and action across a relevant modal subspace, a kind of tracking, if you will. But (G-M) is false. It entails that there cannot be contrast with respect to freedom and orthonomy without a modal contrast. But there can be (section 4).

For those with a keen eye for the developments in other philosophical subdisciplines, this is an all too familiar story. For it has been told (and is being told) about some of the most crucial artefacts of philosophical pride (and shame), such as the notions of "action" and "knowledge" (see Heering 2022a for a brief synopsis of the story). In the subdisciplines concerned with these notions, too, a kind of robust sensitivity to the relevant features was once thought to be characterizable as a causal relationship. And faced with seemingly insurmountable counterexamples, these subdisciplines, too, turned to the alluring charm of modalist principles, only to be met with another set of counterexamples (and, as I hope to show below, counterexamples of the same form as those faced by causalism).

But this bleak result is not the end of the story. For the shape of the right kind of account is recognizable in the failures of causalism and modalism. The reason why there can be freedom and orthonomy contrasts without either causal or modal contrasts is that the notions of freedom and orthonomy track facts about explanation, and facts about explanation are not

reducible to either causal or modal facts. A growing minority accepts this conclusion for other notions in need of a robust tracking of features (for example in epistemology: Faraci 2019, Lutz 2020, Bogardus and Perrin 2022). With the help of their insights, I shall give a characterisation of the kind of explanatory fact suitable to ground freedom (section 5). Very roughly, the *explanationist* view I favour holds that facts about free agency are grounded in facts about *unifying* reasons-explanations of actions.

Before we delve into the metaphysical details, however, it is worth highlighting how the view that explanatory facts are the determiners of freedom is dialectically linked to another part of the free will debate: Frankfurt Cases and so-called “actual sequence” approaches to freedom. I believe that the main insight in Frankfurt’s original 1969 paper is precisely that freedom and responsibility are, at base, explanatory concepts, and that the grounds of freedom should reflect this fact (this is one of the many ways in which Frankfurt’s results were ahead of their time). So understood, Frankfurt Cases lend immediate support to the explanationist view, and correspondingly the explanationist view should be seen as the most direct heir of Frankfurt’s insights properly understood – an “actual sequence” view in the strictest sense. I shall explain these points presently. I will also clarify how my account relates to cases of acting against reasons – where there are seemingly no reasons-explanations to come by.

1. Frankfurt on Freedom and Explanation

Niklaas is considering breaking up with his partner. He can see reasons for and against staying in the relationship and he weighs them carefully. However, he decides that his love for another person is a sufficient reason to break up with his partner, and so he does. Unbeknownst to Niklaas, his rival has implanted a chip in his brain which is monitoring Niklaas’s every mental move. Should Niklaas decide to stay with his partner, the chip will detect this and make Niklaas (decide to) break up with his partner. But Niklaas decides on his own, so the chip remains dormant. This is a Frankfurt Case (Frankfurt 1969).

Niklaas, most of us would think, is free with respect to and responsible for his action. But Niklaas lacks the ability to do anything other than break up with his partner, or so it would appear when first encountering the case. For in any relevant situation in which Niklaas makes some mental moves towards not breaking up, the chip springs into action and course-corrects. Thus, the canonical reading of Frankfurt’s conclusion is that the so-called “ability to do otherwise” is not necessary for freedom and responsibility.

It is a sad fact about Frankfurt Cases that the discussion that followed in their wake has become what Frankfurt himself has called “disconcertingly intricate” (Frankfurt 2003, 339). But this intricacy is created solely by the vast distance that now exists between the complicated state

of dialectical play and Frankfurt's simple original points, which large swaths of the debate seem to be alienated from. Let us therefore call these simple points back to mind:

At the conceptual core of Frankfurt's paper is the distinction between elements that *make it inevitable* that someone does something and elements which *explain* why someone does something. Strictly speaking, Frankfurt Cases only show how these two elements can come apart. In order for S 's ϕ -ing to be inevitable, all that needs to be true is that S *would have ϕ -ed anyway*. Clearly, whatever establishes this counterfactual need not be involved in why S *actually* ϕ -s. Thus, explaining and making inevitable may come apart, even though we tend to overlook this because in typical cases of unfree action what makes the action inevitable is also what explains it.

Because all that Frankfurt Cases show is the distinction between explainers and inevitability makers, the canonical conclusion that the ability to do otherwise is not necessary for freedom and responsibility is not established by them. This conclusion is established only in conjunction with the second major insight of Frankfurt's paper, namely that *only explainers of action are relevant to freedom and responsibility*. In accordance with my focus on the grounds of freedom, we should read this principle as a kind of guideline that tells us what types of elements, in general, are eligible to be grounds of freedom:

(Relevance) Only factors that explain an action can ground facts about freedom with respect to that action.

For now, (Relevance) should strike us as eminently plausible. Imagine that Niklaas tried to claim that he is breaking up with his partner because a mysterious villain has credibly threatened to kill him if he continues the relationship. If this is indeed true, then, tragic as this situation may be, it will offer some protection against some of the blame levelled against Niklaas for the break-up. It will offer protection, moreover, by suggesting that in a sense Niklaas's hands were tied—that his action wasn't free to the full extent. But it will do so only if it is also true that he broke up with his partner *because* he was threatened by a mysterious villain. If his partner has good reason to suspect that the explanatory claim is not true, then they can call bullshit on Niklaas. "Bullshit!" they will exclaim "That's not *why* you are breaking up with me!". Niklaas's partner is cancelling Niklaas's attempt to escape responsibility by putting pressure on the explanatory claim.

Since Frankfurt cases show that the ability to do otherwise is sometimes not what explains why someone did what they did, (Relevance) gives us the result that in these cases the ability to do otherwise is not what grounds free agency. What happens in nearby situations, in other words, is a ground of freedom and responsibility only if it is also explanatory of the actual action. But merely possible affairs don't explain what actually happens. So they never (or very rarely) ground free agency. This reasoning gets complicated, of course, in cases in which it seems all

the available grounds for freedom are potentialities that do not play a role in explaining action—such as cases where agents act *against* their best reasons, but could have acted in accordance with them. I will address this complication below.

If my reconstruction is correct, it already shows that large parts of the literature on Frankfurt Cases are awkwardly discontinuous with Frankfurt's original points. A large contingent of writers have tried to show that the agent in Frankfurt Cases keeps a particular ability to do otherwise, while Frankfurt's supporters have tried to come up with more complex cases in which this ability is missing. But Frankfurt's original point was precisely that Niklaas is free and responsible *whether he has this ability or not*. Thus, even if Niklaas *keeps* the ability to do otherwise, as long as this ability is not in any way involved in explaining his action, it will not be a ground of freedom with respect to that action. We don't really need Frankfurt Cases, then, to get to the conclusion that the ability to do otherwise does not ground freedom. Let us assume, for a moment, that the reason why the involvement of the mysterious villain in the case above can potentially undermine Niklaas' freedom is that he robs him of the ability to do anything other than break up with his partner. His partner will still be right, in that case, that as long as the presence of the villain does not explain his actions, Niklaas has no business bringing them up as an excuse. So now imagine that Niklaas responded by admitting that yes, the villain's threat was not as strong as he made it out to be, and imagine that he apologises on this basis. This is still bullshit. His partner, if attentive, will tell him that he has missed the point. He is apologising on the wrong basis because the fact that he could have refrained from breaking up with his partner plays no role in why he in fact broke up. An apology of the kind Niklaas is trying to give makes sense if it identifies a ground of freedom. No such ground has been identified. The reason cases like this have been largely missed, I believe, is that explanatory claims of the form 'p because q' have at least two ways of being false, and we tend to focus on only one. It can be the case that q is false (or p is false), or it can be the case that both p and q are true, but it is false that q explains p. Frankfurt focussed on the first kind of case in his paper because it offers a less messy way of presenting his point. But he could have given an example of the other kind, and it still would have made exactly the same point.

Frankfurt's argument teaches a further lesson that supports the key intuition behind orthonomy accounts. (Relevance) only contains reference to facts about freedom, including facts about the *lack* of freedom. And there are many ways in which we can explain actions, only some of which will make an action free. Others will undermine freedom. For example, explanations of action that proceed in terms of neurological factors, or indeed explanations in terms of the agent's compulsive states, will typically be taken establish that the agent's action was not free.

Niklaas, on the other hand, can offer an explanation in which his reasons for action play an indispensable part – a reasons-involving explanation of his action. He can account for his action in terms of the normative demands he saw and understood. He can cite these normative

demands as the basis of his behaviour. Moreover, these factors don't just go *some* way towards justifying his action, they are *sufficient* for it.¹

That Niklaas can give a rational explanation of his action in terms of the reasons for which he acted is plainly a key influence of our intuition that Niklaas is free and responsible. Thus, we can derive a positive specification of (Relevance):

- (E) Free agency with respect to S's ϕ -ing is grounded in a reasons-explanation of S's ϕ -ing in terms of sufficient normative reasons.

The idea that Niklaas's action is free because he can explain that action in a particular way fits with the basic tenet of orthonomy views. The central claim these views make is that freedom is grounded in a special relationship—a kind of sensitivity or responsiveness—to reasons. I have said that this orthonomous relationship must consist in a robust acuity to the relevant normative demands. And we can account for this relationship best, or so I shall argue, by holding that it is the kind of relationship that is expressed in reasons-explanations of action.

Before I continue, however, an essential dialectical caveat concerning the scope of my claims in this paper needs to be addressed.

2. A Caveat Concerning 'Negative Cases'

My claims in this paper will be restricted in an important respect, because they will, perhaps unusually, not concern what I shall call "negative cases". Let me explain.

Accounts that seek to ground freedom in orthonomy face a general structural problem: Freedom distributes symmetrically across "positive" cases of acting for reasons and "negative" cases of acting against reasons. That is, there are clearly cases of freely acting *against* sufficient normative reasons. Assume for example, that I come across an injured person on my way to an important job interview for a job I don't depend on but want. I know I should help, but I also know helping would mean missing the interview. So I decide not to help the person. In refusing help, it seems, I am acting freely – and many would point out that we base our subsequent attributions of responsibility and blame on this assessment. Orthonomy, however, does not distribute symmetrically across positive and negative cases, or so it would appear. That is, only positive cases of acting for reasons would seem to be cases of acting orthonomously, i.e. cases in which the orthonomous connection between reason and action obtains. For in the case of the injured person, my decision to refuse help is clearly not responsive to the normative balance of the situation. This mismatch also infects (Relevance) above, for in cases of acting against sufficient reasons something other than what is explanatory of the actual action-

¹ We should adopt a weak reading of sufficiency here, according to which an agent in buridans ass style situations has two sufficient reasons to act.

perhaps a missed opportunity or unexercised ability—appears to be a determiner of freedom.² What makes my driving on to the job interview free, in other words, seems to be that I have the ability or opportunity to stay and help the injured person (i.e. my ability to act for sufficient reasons), not the elements that actually explain my action.

Accounts of freedom as grounded in orthonomy owe us a solution to this structural problem. The default solution in the literature is to fall back on a “capacitarian” notion of orthonomy, according to which the mere possession of reasons-responsiveness is enough to ground freedom. This option reestablishes symmetry: in both good and bad cases, what is required for freedom and orthonomy is the possession of the ability to respond to reasons. But it comes at the cost of partially abandoning the strongest intuitions in favour of adopting an orthonomy view in the first place. For as we saw, at least part of the motivation to adopt an orthonomy view of freedom derives from thinking about cases in which agents actually manifest acuity to the normative situation—cases in which they act *for* or *on the basis of* reasons, that is. Someone who merely possesses the ability to respond to reasons and who does what those reasons recommend may still not be responsive in this sense. For they still might just accidentally do what their reasons recommend, thus manifesting no relevant acuity to the normative situation (indeed, this idea about accidentality will be central to this paper).³ What positive cases of acting for reasons seem to require, then, is the exercise, not merely the possession of the ability to respond to reasons.⁴ The same set of intuitions is at work in Frankfurt’s appeal to (Relevance) discussed above and the Frankfurt-Cases with which it interacts. In order for (Relevance) to even apply to Frankfurt-Cases, they need to feature agents who not only possess the relevant abilities, but exercise them too. For unexercised abilities to x might have no relevance to why someone x-ed.⁵

² Cases like these accordingly play a role in discussions surrounding (Relevance), for example in Ginet 1996, Widerker 2000 and 2003, Palmer 2014 as well as Swenson 2015. There are also cases in which facts about certain obstacles to the successful performance of an action seem to matter to judgments about freedom such as the SHARKS case featured in Fischer and Ravizza 1998. See also Sartorio 2011, 2016.

³ The idea that an agent’s successfully x-ing is attributable to them only when they exercise(d) the ability to x is most thoroughly explored in virtue epistemology, where successes attributable in this special sense are called ‘apt’, while successes in which the agent’s abilities are uninvolved in this sense are called “accurate” (for the classic treatment, see Sosa 2007: 22-3).

Extant orthonomy approaches often oscillate between descriptions that invoke capacitarian intuitions and descriptions that invoke exercise intuitions. For example, here is how Susan Wolf presents her view:

The present view, in contrast, denies that responsibility rests on the availability to the agent of at least two options. What matters is rather the availability of one very particular option, namely, the option to act in accordance with Reason. If, on this view, the agent exercises this option, then it is irrelevant whether the agent might not have exercised it. In other words, it is irrelevant whether the agent had the ability to act in discordance with Reason. If, on the other hand, the agent does not exercise this option, then the question of whether she could have exercised it is all-important. (Wolf 1990, 68)

The first two sentences, with their focus on the mere availability of options, suggest that what is important to Wolf is the mere possession of an orthonomy-ability (and indeed, this is what her official view is). But the rest of the paragraph is phrased under the condition that one of these options (the one leading to positive cases) is chosen, and so suggests that what really matters is not only having an option, but following (or rather having followed) through with it.

⁵ Hu (2023) discusses senses in which unexercised abilities might be explanatorily relevant. A subplot of this paper will be that insofar as these senses reduce to modal or causal readings, they are not of the right kind to ground freedom.

If we wanted to honour both the intuitions springing from bad cases—which seem to favour a capacitarian approach—and the intuitions springing from good cases—which seem to favour an exercise-based approach—, we might be inclined to drop our desire for symmetry. We might hold, that is, that in bad cases, freedom and orthonomy require the possession of the relevant abilities, while in good cases, they additionally require the exercise of these abilities. My own allegiances lie with a more controversial third option, which accepts symmetry and holds that both in good and in bad cases, the exercise of abilities is required. Thus, my full picture of orthonomy that I will only be able to hint at in this paper will centrally rely on a notion of exercising abilities badly, and it will offer an account of how even in cases of acting against reasons, those reasons are not irrelevant to the explanation of action.⁶

The same solution will account for a second type of negative case: The negative cases covered directly above involve an agent acting against what they know to be a sufficient reason. In another kind of negative case, agents falsely believe some consideration to be their sufficient reason for x-ing and fail to y for sufficient reasons as a result. It is an open and controversial question in the philosophy of action what we ought to say about such agents⁷, a question I cannot hope to do full justice here. However, part of the full picture of freedom grounded in orthonomy will be an account of how agents in less than epistemically ideal circumstances still exercise their ability to respond to reasons (and can thus be counted as free).

But since the development of this full picture lies well outside the scope of this paper, I will here want to restrict my claims only to the positive cases. That is, what I am concerned with here is finding the right account of the kind of orthonomous relationship that obtains in the positive cases, irrespective of whether the same conditions ground freedom in the negative cases. All my claims should be understood with this restriction in mind. In particular, my objections to the causalist and the modalist project are that they cannot capture what happens in the positive cases, irrespective of whether their ability to capture the negative cases gives us independent reason to stick with those projects. Since it is primarily negative cases in which the notion of “exercising abilities” becomes especially pertinent, moreover, I will omit this notion from my positive explanationist account in section 5 and the discussion preceding it. My positive account will be compatible with approaches like that of Neta (2019) and Marcus (2012) that can be interpreted as spelling out the orthonomous relationship in terms of the exercise of rational abilities. But my ambition here is to highlight and describe the particular

⁶ This full picture will offer an account of how we explain immoral (or more generally irrational) actions. Frankfurt himself suggests this route in Frankfurt 2003. I offer a general argument for why the notion of exercising abilities badly is required anyway in Heering forthcoming

⁷ In particular, it is an open question whether there is a sense in which the “agent acts for a reason” given there is no reason the agent acts for. Some accordingly hold that agents with false beliefs about their reasons act “for no reason at all” (Alvarez 2018, Littlejohn 2012), while others introduce “apparent” or “subjective” reasons, which are also genuinely part of the normative situation (see Schroeder 2008, Comesaña and McGrath 2014).

type of explanatory profile that facts about orthonomy require, irrespective of whether the same facts can be described via the notion of “exercising rational abilities”.

With this out of the way, I will now present some reasons to doubt the causalist and modalist accounts of freedom-grounding orthonomy.

3. Causalism about Freedom and Orthonomy

Carolina Sartorio (2016) has recently offered a powerful new “actual sequence” view supported by a sophisticated metaphysics of causation. She follows the conclusions of Frankfurt Cases, but believes they are better captured in terms of causes:

(G-C) Freedom facts are grounded in causal facts.

By “causal facts” Sartorio simply means facts of the type “X caused Y”.

Notably, Sartorio is also dedicated to the idea that Niklaas’ positive freedom status is grounded in his sensitivity to reasons, which she thinks are “the right kind of causes”. Her way of capturing the orthonomous relationship between reason and action—the robust rational acuity required—is to claim that the agent is sensitive to the *absence of counter-reasons*. Sartorio has a well-worked out view on absence causation, which she can just plug in at this stage.

The absence-causal view does capture something crucial about orthonomous relationships. Sensitivity to a demand seems to entail that the agent would not just blindly act in the same way no matter what. After all, if that was the case, we might have reason to suspect that the agent’s action matching with the demands of their reasons in the actual case is a mere coincidence, a case of the normative world accidentally lining up with what the agent was going to do anyway. In order to avoid a modalist interpretation of this intuition (“Niklaas would have acted differently, had his normative situation changed”), which Sartorio thinks violates Frankfurt’s insights, Sartorio opts for the absence-causal reading.

Sartorio describes her position as a causal theory “par excellence”, which is to say that it holds that “[...] actual causal histories are all that matters” (Sartorio 2022, 1255) and she aligns herself with other infamous causalist proposals in the history of philosophy, like Goldman’s theory of knowledge (Goldman 1967) and Davidson’s theory of action (Davidson 1963). Both notions, notably, require an orthonomous connection to the relevant parts of the world.

The problem with this is that causal theories par excellence in general have been plagued by a potentially lethal problem, which also befalls a causalist treatment of freedom and orthonomy. This is the problem of so-called causal deviance. It comes in the guise of a set of examples that directly falsify (G-C).

To see this, compare the following two cases:

Freedom Contrast with no Causal Contrast

David is the member of a strange religious cult with strict rules regulating all conduct. Among many other things, David's god decrees, on the basis of obscure metaphysical principles, that you ought to break up with your partner if you love someone else. And David does love someone else. So through his god's command he breaks up with his partner because he loves someone else.

Adina is a member of the same cult, and she also loves someone else. She also knows that her god decrees that love for someone else, through obscure metaphysical principles, must lead to break-up with your current partner. But through some felicitous stroke of sight, Adina can also see that according to the rules of monogamous relationships it would be unfair to stay in the relationship while loving someone else. Through this recognition, she breaks up with her partner because she loves someone else.

I want to make two related claims about these cases.

First, they represent a comparative situation, in which there is a *difference in free agency without there being a difference in the relevant actual causal sequence*. Thus, they falsify (G-C). This is because David is not free and responsible with respect to his action—or at the very least he is less free and responsible than Adina. David, after all, is very much like a deluded agent, or the irredeemable dictator Jojo, who is treated in the literature on orthonomy as the prime example of a non-orthonomous agent (see Wolf 1987). He is cut off from normative reality, perhaps permanently unable to recognize many of its demands. The rules of his god are not a secure guide to this reality (on the contrary, they are a kind of insulation from it), even if in this case, they happen to overlap with the demands of normative reasons. Adina, on the other hand, does not seem equally unhinged. She recognizes the demands of her situation and acts on their basis, even if it is surprising that her cult membership has not had the familiar distorting effect on her ability to see and act for reasons. Her action, therefore, is free and responsible.

The crucial point about comparing Adina and David is that all the relevant *causal* facts are the same for them. After all, both of their actions are caused by the same fact, the fact that they are in love with someone else.⁸ What is different between David and Adina are the *ways* in which they arrive at the belief that this fact is a sufficient normative reason to break up with their partners. They have different grasps of what *makes* the fact that they love someone else a sufficient normative reason to break up. This is primarily a difference about the non-causal

⁸ In the following, I assume factualism about the reason for which agents act, i.e. the view that facts, not mental states, are the reasons for which agents act. I take to be the more plausible view, both in general and within an orthonomist framework. The arguments can be reproduced in a psychologist framework.

normative explanation for why the fact that they love someone else should matter causally. David and Adina, in other words, differ with respect to the normative grounds of their reasons. David and Adina might also differ with respect to the causal processes which link their bases of action and their actions. But not just any different causal process will be relevant, for it might still be a causal process that grounds the fact of the form "R caused A" (where R designates a reason and A designates an action).

The causalist seeks to understand the orthonomous relationship between reason and action in terms of causal facts of the form "R caused A".⁹ I have presented an example in which this fact remains the same between two agents while they differ with respect to their freedom. One way for the causalist to contest this is to be more precise about what the agent's reason is. A natural thought about David and Adina's case is that that fact they are in love with someone else is not their *complete* reason for breaking up. Surely, the thought goes, what David and Adina believe about the rules of their god or morality must be part of their reason as well. So, for example, perhaps a complete description of David's reason is that he is in love with someone else *and* that the rules of his god favour a decision to break up-or something along those lines.

But this strategy will only lead the causalist into more trouble. For one thing, it just isn't true that agents always need to have beliefs about the validity of their inference rules to reach conclusions in practical inference. Most agents simply manifest certain competences or dispositions by moving directly from premises to conclusions. The contents upon which they perform these operations are reasons-in David and Adina's case, it would be that they love someone else; the relevant rules of inference don't show up as extra premises. In fact, if we have learned anything from Achilles and the Tortoise (Carroll 1895), it is that rules of inference *cannot* show up as extra premises on pains of infinite regress. So the idea that whatever is responsible for the deviation in the deviance case above will just be part of the relevant reason is doomed to fail. For what makes the process between reason and action deviant may be a background condition-like what practical inference rule you are committed to. And if that is so, baking the deviance-maker into the very reason is tantamount to collapsing the distinction between background conditions and reasons for action.¹⁰

Hence, David and Adina may differ in background conditions or even causal processes, but not in terms of facts about what caused their actions. This is a general feature about cases of

⁹ There might of course be a weaker version of causalism according to which there is no difference in freedom without a global causal difference in the situation. This version would have abandoned the idea, however, that there is something unique about sensitivity to reasons, that reasons are the "right kind of cause", which is why I won't comment on it here.

¹⁰ See Lord and Sylvan 2019, 150 for a similar argument

causal deviance: they work with the insight that facts of the form “X caused Y” are unspecific in that they may have several verifiers depending on what specific causal process links X and Y.¹¹

Within Sartorio’s framework it is also important to point out that David and Adina’s situations are the same with respect to all absence causes. According to Sartorio, Adina’s action is also caused by the absence of overriding counter-reasons. For example, Adina’s action is caused by the absence of the fact that breaking up would ruin a beautiful relationship. But we need only imagine that David’s god has decreed a system of rules that—through mysterious metaphysical principles—matches with the patterns in which these absence causes work. For example, David’s god might have decreed—through strange religious rules—that a beautiful relationship is an overriding reason to stay with a partner. In fact, this will make David even less susceptible to his normative situation because it will move him closer to an agent globally deceived about the real reasons to do things, an agent who does everything correctly accidentally.

It is now commonplace to avoid eye-contact with the magnitude of the problem of deviance by adding the “in the right way” clause to your allegedly purely causal account of x, where “causing in the right way” is either supposed to be a placeholder for the metaphysical work I suggested above is unlikely to be forthcoming anytime soon, or it denotes a primitive notion. Sartorio also relies on the phrase in her account, so we might suspect that she will say that there is after all a difference in the causal facts between David and Adina: Adina’s situation includes the following fact: the fact that she loves someone else causes her action *in the right way*.

But in the context of the attempt to ground freedom in orthonomy, we cannot rely on an unanalysed notion of “in the right way”. For, as the contrastive case of Adina and David shows, since “in the right way” describes precisely the element that makes the difference between an orthonomous relationship and a relationship that falls short of orthonomy, it also draws the line between those causal structures that ground freedom and those that don’t. And surely any causalist account of freedom must be able to say something informative about the difference between when agents are free and when they aren’t. Providing an informative answer to the problem of deviance, in other words, is one of the primary tasks of any orthonomy view of freedom.

This brings me to my second point. The problem of deviance, most abstractly conceived, is a problem of accidentality (Aguilar 2012).¹² David’s action isn’t orthonomously related to his reasons despite their causal connection because David *accidentally* does what his reasons

¹¹ The same principle applies to the adjacent idea that David and Adina’s actions differ in terms of at least one of their causal ancestors: the way in which X causes Y does not obviously reduce to the involvement of an intermediate or prior cause Z.

¹² The theme of the problem of deviance as a problem of (a type of) luck is also central to the work of Ernest Sosa, who not only tells a similar story about causalist approaches to orthonomy notions (see for example Sosa 2017, ch.8), but also offers a similar solution in terms of the exercise of capacities.

demand. It is for this reason that it is hard to see how we can ever account for the difference between deviant and non-deviant cases in *purely* causal terms.¹³ It simply isn't plausible that accidentality is a phenomenon that plays out on the level of causes.¹⁴ Whether or not we count some match as accidental, as I shall elaborate below, is essentially a matter of how we *explain* it. Accordingly, key to a solution, I believe, is to recognise the difference between causation and explanation. Quite generally, causal relations—singular extensional relations between particulars—don't 'filter up' in Lando's (2017) words, to the important explanatory relations—general, (hyper)intensional relations between facts. There can be a difference in explanatory facts, in other words, without there being a difference in causal facts. And this is precisely the case in the David-Adina example. Even though his loving someone else causes David's action, it is not the case that David breaks up with his partner *for the reason* that he loves someone else. For David does not recognize this reason under its proper guise. There are of course, explanatory relations which relate 1:1 to their underlying causal structures, because we can always adopt views on either causation or causal explanation in which one is assimilated to the other. There is therefore a thin causal sense in which the fact that he loves someone else does explain David's action. But the crucial point is that this sense does not amount to a reasons-explanation of action. It does not amount to the kind of explanation that we can give of Adina's break-up, that is. And it is this "thick" kind of explanation that seems to be the crucial difference as far as Adina's free agency and David's lack of free agency are concerned.

This second observation suggests that any orthonomy view of freedom would do well to account for the difference between Adina and David in explanatory terms, and indeed this is what I shall suggest below. But my proposal will make more sense if we first consider the second extant orthonomy view of freedom, which, on the face of it, can *also* account for the difference between David and Adina. This account surmises that what grounds freedom aren't causal facts, but modal facts about the agent.

4. Modalism about Freedom and Orthonomy

Perhaps the best-known extant accounts of freedom and orthonomy are modalist (Smith 2003, 2004; Vihvelin 2004; Vihvelin 2013; Mckenna 2013). They ground freedom in facts about what would have happened had certain conditions obtained, or in what happens in relevant sets of alternative possibilities. They hold:

¹³ There are of course many notions of causation out there that are already assimilated to the kind of explanatory function I have in mind, such as Eric Marcus' "rational causation" (Marcus 2012) or perhaps versions of aspect causation that works with normative aspects. I don't consider these notions the right material for the causalist par excellence, because they are loaded with irreducibly explanatory ideology already.

¹⁴ Although traditional causal models appear in Owens (1992), Monod (1970), Horwich (1982).

(G-M) Freedom facts are grounded in modal facts.

The draw of such approaches is that they seem to be especially well-suited to account of the kind of sensitivity to reasons that orthonomy views are after. For their idea is that the relevant sensitivity can be expressed modally. An agent who is sensitive to the demands of their reasons, so the idea goes, would change their action accordingly if those demands changed as well – if they were overridden or strengthened. For example, if Niklaas realised that his love for someone else was just a temporary folly, or if he realised it was much stronger than he initially thought, he would presumably adjust his actions—he would not break up with his partner, or break up with his partner even more quickly, respectively.

Notably, modal truths also feature as a longstanding and prima facie promising candidate for getting a grip on the distinction between David and Adina. That is, modal truths have been used to describe what goes wrong in cases of causal deviance (see Morton 1975; Bishop 1989, 150; Shope 1992, 256 ff). The idea is this: If David's action is connected to the fact that provides a reason for him through the strange demands of his god, then no matter how we wiggle the situation, he will always break up with his partner. For example, even if his love for someone else is a fleeting folly, his god will, through strange metaphysical principles, still require David to break up. Thus, David's action is not sensitive to the demand of his reasons, because his action does not properly track that demand—its waning and waxing—through modal space. What this dialectical connection shows is that both causalism and modalism are tracking an intuition about non-accidentality: the worry is that the fact that agent does what is recommended by their reasons alone does not establish an orthonomous relationship. They might have done what their reasons recommend *accidentally*, after all.¹⁵ Thus, to establish the needed non-accidentality, we need a kind of alertness to shifting normative situations—either in terms of the causal powers of absent counterreasons (causalism), or in terms of what would happen if these reasons were present (modalism).

The problem is that modal truths are also not enough to establish orthonomy and freedom, and as I hope to show, this is for reasons that are interestingly systematically related to the failure of causalism.

Again, start with a pair of cases.

Freedom Contrast with No Modal Contrast

Roland has severe trouble seeing what exactly his reasons are for anything. He has a general concept of what he ought to do, but he regularly fails to correctly apply it to the particulars of situations. He regularly fails to identify the right facts bearing on what he ought to do, and if he does, he misjudges both the weight of reasons and what

¹⁵ See McKenna 2013, 154 for a nice illustration of the worry.

options they favour. Fortunately, Roland has been fitted with a novel mind-attunement device, which selectively links his mind with that of his friend Daniele. The device allows Daniele to subconsciously monitor the situations Roland finds himself in. If a normatively important situation arises, Daniele will notice and then transmit what he would do if he found himself in that situation. From people generally recognized as authorities in his community, Roland knows that Daniele is extremely good at responding to reasons. So he relies on the prompts from the device to make his decisions. One day, Roland is faced with the decision to break up with his partner and receives the information that if Daniele was in his situation, Daniele would break up with his partner. Consequently, Roland breaks up with his partner.

Daniele is very good at seeing, understanding and translating into action reasons for actions. While pondering his relationship, he realises that the fact that he loves someone else is a sufficient reason to break up with his partner, and so he does.

Let me again make two claims about these cases, before I briefly consider a modalist response.

First, the two cases represent a situation in which *there is a freedom contrast but no modal contrast*, falsifying (G-M). This is because Roland's action of breaking up with his partner isn't free, or at the very least it isn't as free as Daniele's action is. The impression that Roland does not act freely in breaking up with his partner is backed up by the fact that Roland does not act autonomously. He is merely imitating Daniele, after all, without any real understanding of why he is doing the things he is doing. He wouldn't be able to account for his conduct, other than shrugging his shoulders and deferring to Daniele. He is a blind follower, which is the paradigm of an agent whose actions are heteronomously determined. As above, this impression is in lock-step with the intuition that Roland does not act for the reason that he loves someone else. Roland accidentally does what his reasons recommend, but he does not act *for* those reasons. This is, importantly, despite the fact that his actions are modally harmonised with his reasons. Roland has the disposition to do what Daniele does, and Daniele, being the highly skilled responder to reasons he is, would course-correct his actions should his reasons wax or wane. If Daniele found out that his love for someone else was only a temporary folly, he would not break up with his partner. And because Roland is so good at knowing and doing what Daniele would do, he wouldn't either. Thus, the modal profiles of Roland and Daniele are exactly alike, while only Daniele acts freely.¹⁶

Second, even though there is no contrast in their modal profiles, there is a notable difference in how we can *explain* Daniele's and Roland actions. Daniele can say that he broke up with his partner because he is in love with someone else. Roland has no such answer available. All he

¹⁶ If you are worried that Roland is an unfair case to field against modalism because his reasons don't cause his actions, note that we could simply run the argument with a modally enhanced version of David the religious fanatic.

can say is that he did it because Daniele would, and Daniele is pretty good at doing the right thing.

It is perhaps helpful to point out that the type of modal accidentality that Roland exhibits is the result of a more general feature of modal truths, which is that they can mimic features of the actual situation. A rock rigged to an explosive device which shatters the rock every time it hits a hard object modally mimics fragility, even though the rock, in a sense, isn't actually fragile. Roland's disposition to imitate Daniele mimics the modal profile of a responsive agent, even though Roland is anything but.

Some extant modalist accounts of freedom and orthonomy - sometimes referred to under the label "New Dispositionalism" - are designed to counteract issues to do with modal mimicking and its twin, modal masking (Fara 2008, Vihvelin 2004, 2013, Smith 2003, 2004). As they would have it, there is a modal contrast between Roland and Daniele as long as we latch onto the right counterfactuals. The idea is that we get cases of masking and mimicking only by introducing elements that are interferential to the normal functioning of agents. In assessing which sets of alternative possibilities are relevant, we are allowed to "abstract away", as it were, from these elements. For example, in assessing whether the rock in the above example is fragile or sturdy, we are allowed to abstract away from the explosive device and look at the rock as it is intrinsically. We will then assess the modal profile of the rock with respect to breaking and hitting hard objects in terms of alternative possibilities in which the device is absent. Since the rock remains intact in those situations, there is a sense in which it remains sturdy.

An advocate of this strategy might insist that there is a modal difference between Roland and Daniele, because the influence of Daniele is extrinsic to Roland's agentic systems and can thus be "abstracted away". In situations in which Roland's reasons slightly change and Daniele is absent, Roland fails to properly respond. Thus, if we compare Daniele and Roland-without-Danieles-influence, we do get a modal difference that accords with our judgments concerning their freedom and orthonomy.

There is a larger looming question here about what sense of orthonomy remains when all interferential factors are abstracted away from, and whether this sense is robust enough to ground freedom (Clarke 2008; Whittle 2010; Franklin 2011, Heering 2022b). In this paper, I merely want to make the more basic point that the abstracting away strategy will not render coherent ways to identify modal contrasts for all cases. The trouble is that the distinction between extrinsic and intrinsic elements does not consistently track our notion of when agents remain reasons-responsive in the relevant sense. Imagine that Daniele is only so good at responding to reasons because he (secretly) relies on a wise friend, who explains all relevant considerations to Daniele. It seems to me that this addition shouldn't change our assessment

of the case. But if the New Dispositionalists abstracts away from Daniele's influence on Roland, then they should, by the same token, abstract away from the influence of the wise friend. And this will give us the wrong result. For Daniele-without-his-friend is just as bad, or even worse, at responding to reasons as Roland-without-Daniele. The only way for the New Dispositionalist to avoid this conclusion is to measure with a double standard – that is, to compare Roland-without-Daniele to Daniele-with-his-friend. There is nothing in the New Dispositionalist framework that could justify this double standard.

These additions to the cases of Daniele and Roland will not be new for those who followed the history of the problem of deviance. For cases structurally like that of Roland and Daniele also became pertinent in the literature on deviance, where they accompanied the disappointment over the unredeemed hopes of modalism. Here is Peacocke, poignantly expressing his disappointment with modalist conditions for the notion of perception:

I conjecture that for any appropriate noncircular counterfactual, however complex, we can always find some imaginary example in which the holding of that counterfactual is either not necessary or not sufficient for perception; and this is true, if it is, because whether someone's experience is a perception or not is a matter of how things are actually related to him, not a matter of how they might be. (Peacocke 1979, 136)

What Peacocke expresses here is not just doubt about any particular modalist account, but the very possibility of a modalist account par excellence. This situation bears a striking structural similarity to the problem for causal theories par excellence of orthonomy. Just like it seems unlikely that we will be able to find a purely *causal* difference between orthonomous and non-orthonomous relationships, it seems unlikely that a purely *modal* difference will be found. This is because it doesn't seem like accidentality is a purely modal phenomenon. Accidentality is essentially a matter of how we explain things. And it is notoriously difficult to reduce explanatory facts (of the "thick" sort I've been focussing on) to modal facts, not least because the former are hyperintensional.¹⁷ The solution is therefore again to recognize that modal and explanatory facts come apart.

We should develop a theory of freedom and orthonomy which accounts for the fact that freedom and orthonomy are sensitive to explanatory judgments. The second part of this paper will do exactly that. It will develop an explanationist account of the orthonomous relationship that grounds freedom. Let me summarise first, however. There are two things we can learn from the failures of causalism and modalism.

One: there can be a change with respect to facts about the freedom of an action without there being a change with respect to either the causal or the modal facts. But there cannot be-at

¹⁷ The proposition that A explains B is hyperintensional iff A or B cannot be replaced by intensionally equivalent contents *salva veritate*.

least as far as my cases show—a change with respect to facts about freedom without a corresponding change with respect to facts about the explanation of action. This result lends support to what I argued a Frankfurtian position should claim anyway:

- (E) Free agency with respect to S's ϕ -ing is grounded in a reasons-explanation of S's ϕ -ing in terms of sufficient normative reasons.

The twist behind my examples is that a change in explanatory facts will not necessarily lead to a change in causal or modal facts. Since our intuitions about freedom and orthonomy seem to track facts about explanation, we can create the no contrast cases above by changing facts about freedom and explanation while leaving the causal and modal facts untouched.

Two: there is another notable contrast between agents in the above cases that doesn't track causal or modal facts. In the examples, free agents non-accidentally do what their reasons demand, unfree agents accidentally do what their reasons demand. We also saw that both the causalist and the modalist models are driven by intuitions about non-accidentality. In fact, the special feature of orthonomous relationships I called "robust rational acuity" for lack of a better term might be more accurately described as a non-accidental connection between reason and action.

My idea will be that that we can illuminate (E) by focussing on the relationship between explanation and non-accidentality. More precisely, the account rests on the insight that explanations that ground freedom are *accidentality-dispersing*.

5. Explanation and Non-accidentality

A key intuition I identified in accounts that seek to ground freedom in an agent's orthonomy—their relationship to the normative demands of their reasons—is that orthonomous relationships are non-accidental relationships between reasons and actions. Intuitions about accidentality are a species of intuitions about coincidence, which are ubiquitous in human lives, not only in our practice of ascribing agency. Consider: This pancake looks like Elvis Presley, what a coincidence. Amber's hair-do looks exactly like that of Greta Garbo, what a coincidence. And even though they seem to express statements, not questions, we know from experience that they usually occur within our general endeavour to make sense of the world—to understand, and therefore to explain certain facts, that is. Coincidences attract our attention because they involve certain salient matching relations—the matching between Elvis and a pancake, between Amber's and Greta Garbo's hair-do—that clamour for explanation.¹⁸

¹⁸ Coincidence in its most general form might also concern larger pattern of co-occurring facts. And coincidence-questions sometimes also concern mis-matchings. I leave that to the side here.

I propose that we think about the notion of coincidence as determined by our explanatory practice. When we notice the relevant facts about salient matchings, we can ask why-questions about them, which I shall call *coincidence-questions*. Our sense of whether some matching-fact is a coincidence or not is determined by what answers we can give to these coincidence-questions. Some of our answers disperse coincidence. Other don't. The key to understanding non-accidental and non-coincidental is figuring what the difference consists in.

Why is it a *mere coincidence*, then, that a particular pancake has the same shape as Elvis Presley's head? The core of our intuitions about coincidence is that there is nothing connecting two matching properties, that they are *independent*, as it were. It is a coincidence that this pancake looks like Elvis, in other words, because the fact that the pancake has an Elvis-headshape is independent of the fact that The King has an Elvis-headshape. More precisely, since we are engaged in explanatory practice when we ask coincidence questions, we can say that a fact about a salient matching is a coincidence if *we can explain its components independently*.

This is a promising approximation. For we can explain my other example via this condition. Why is the fact that Amber's hair-do matched with that of Greta Garbo a coincidence, when it is? Because the explanation we can give of Amber's hairstyle is independent of the explanation we can give of Greta's hairstyle. For example, maybe Amber just likes 50's hair-styles, and she happened to pick one that Greta Garbo also picked many years earlier.

We can also explain why and when some matching is not a coincidence. Assume that Amber is a huge Greta Garbo fan. She explicitly modelled her hair after the looks of Greta Garbo. Now, the fact that Amber has a certain hair-do and the fact that Greta Garbo has a certain hair-do are not explanatorily independent.

It is worth making these results more precise. In doing so, I will partially follow work on coincidence by Tamar Lando, Harjit Bhogal, and Marc Lange. What we are asking about in coincidence questions, I said, is a fact about a salient match between two properties. It is important to distinguish two facts here with respect to their logical form (see Lando 2017, Bhogal 2020). The first is the matching fact proper, which is a *relational fact* about the matching between two facts p and q , such that they both contain some salient commonality. For example, this would be the fact that both Amber and Greta have the same hairstyle, or that both the pancake and Elvis' head have the same shape. Let's express this fact as $[p-q]$.

The second is a fact *entails* the matching fact. It is the composite fact that a has a certain property P and that b has property P . For example, it is the composite fact that Amber has a 50's hairstyle and that Greta has a 50's hairstyle. Because this fact is just a conjunction of two separate independent facts, let's abbreviate it as $[p\&q]$.

We can develop a more precise account of what happens in the above examples by recognising that *explaining why [p&q] does not entail explaining why [p-q]*. This is because we can always explain why a conjunction holds by stringing together two independent explanations. For example: Why does the pancake have the shape of Elvis Presley *and* Elvis Presley has the shape of Elvis Presley? Well, because I dropped the pan and it made that shape *and* because the genetic code of Elvis' parents led to that shape. This is a case in which we can explain the components of a matching fact independently, i.e. a case in which we can explain [p&q] by explaining p and explaining q, but we can't explain [p-q]. Mere coincidences are instances in which we can only give such *composite answers* to the coincidence question, or, if you prefer, when we fail to explain what the question was really asking for, namely why [p-q].¹⁹

There are two formal features about the difference between explaining [p&q] and explaining [p-q] that further illuminate it:

First, when we merely explain [p&q], explanation is what I shall call *conjunctively restricted*. Let X be the explanation for p and Y be the explanation for q. Then, in an explanation of [p&q], X explains p as well as X&Y does, and it doesn't explain q at all. Y explains q as well as X&Y does, and it doesn't explain p at all (see Faraci 2019, Lange 2010; 2016). Take the explanation for why Amber's hairstyle is like Greta Garbo's hairstyle (when this is coincidental): the fact that Amber likes 50's hairstyles doesn't explain why Greta Garbo has the hairstyle she does. And the conjunctive fact that Amber likes 50's hairstyles and Greta's hairstylist picked a certain hairstyle explains the fact that Amber has a 50's hairstyle just as well as the fact about her preferences alone does (the second conjunct does no additional explanatory work).

Second, and relatedly, it is characteristic of answers to coincidence questions which *do* give an answer as to why [p-q] that the "because" operator in them does not function distributively. We are familiar with operators that exhibit this behaviour, perhaps, from deontic concepts. It does not follow from the fact that it is permissible to ϕ and the fact that it is permissible to ψ that it is also permissible to $[\phi-\psi]$. This is because the fact that an agent [ψ -s and ϕ -s] may come with additional deontic constraints due to the interaction - that is, relation - between ϕ -ing and ψ -ing. Likewise, it does not follow from the fact that X explains p and the fact that X explains q that X explains [p-q].

¹⁹ Erasmus Mayr has suggested to me an alternative way of characterising how unifying and non-unifying explanations differ with respect to the relational fact. The idea is that even non-unifying explanations can be successful at explaining the relational fact, but they necessarily require appeal to the matching fact in order to do so. Unifying explanations, by contrast, are those that explain the relational fact directly, without any need to appeal to the matching fact. Since I lack the space here to fully compare this proposal to my own account, I will leave it at the readers discretion which piece of theory they prefer.

I will call explanations with these features, which they inherit from their special explananda [p-q] *unifying explanations*. My results so far can also be summarised in the slogan that *it is a mere coincidence that [p-q] if there is no unifying explanation of [p-q]*.

As we saw, there are different ways to tell a story about the relationship between p and q such that it will not be a *mere coincidence* that [p-q] (think again about Amber's infatuation with Greta Garbo). Just what kind of story we will need to tell, and just how strong a connection between p and q needs to be established, will be a contextual matter. It depends on what kind of explanatory interests are at play. For example, let us say that Amber is not infatuated with Greta Garbo, but with a hairstylist from the 50s. Unbeknownst to Amber, Greta Garbo was also very impressed by this hairstylist. Now, is it a *mere coincidence* that they have the same hair-do? It depends, I think. Some of our explanatory interests about Amber will be satisfied by this kind of "common explainer" answer. But we just as often screen-off such root explanatory factors in our everyday practices, because we are looking for more direct and proximal connections between Amber's and Greta Garbo's hairstyles. There is therefore a plethora of complicated ways in which it might be true that something isn't a *mere coincidence*, which depend on the precise explanatory interests at play and the precise way in which these interests are fully or partially satisfied by our answers. It is outside the scope of this paper to attempt to systematise these senses, nor indeed do I think it is a philosophically worthwhile task. I want to instead focus on the special sense of lack of coincidence expressed in judgements of non-accidentality. For as shall become clear presently, in order for some connection to be *non-accidental*, it can't just be true that it isn't a *mere coincidence*. Non-accidentality requires a particularly strong connection between two facts, and therefore a specific type of answer to the coincidence question.

Accidentality, as I use it, pertains to the notion of coincidence in the context of the ascription of mental states and rational agency. Still, the model just developed applies: Whether the connection between reason and action is accidental depends on how we answer the corresponding coincidence question. It will be helpful to give a clear account of the form of that question in the context of asking about the matching between reason and action.

What we are interested in when it comes to orthonomy is why what the agent does satisfies the demands of their reasons in the situation. That is, in the particular situation it would be rational that the agent ϕ -s (I will abbreviate this as R) and the agent also ϕ -s. What we want to know is whether this matching is accidental or not. Just like before, we need distinguish two explananda in the vicinity: The conjunctive fact *that it would be rational to ϕ and S ϕ -s* on the one hand and the relational fact *that S rationally ϕ -s* on the other. Answering the coincidence question, again, requires providing an answer to [ϕ -R], the fact that S rationally ϕ -s, not just to [ϕ & R]. However, the standards for the kind of unifying story we must tell to provide a successful answer to this question are uniquely high for our ascriptions of rational agency.

In order to see this, we can go through ways in which we can fail to answer the coincidence question, which, not coincidentally (pun inevitable), line up with the history of orthonomy notions, and with sections 3 and 4 of this paper.

First, consider cases of “mere coincidence” for the relationship between reason and action: A roof tile blown from a nearby house hits my head, causing me, extraordinarily, to form the proximate intention to dive into a river and save a drowning child. As it happens, I was hit near a river, in which there is indeed a child in urgent need of saving. I execute my intention. In this case, I have no interest in the moral demands in the vicinity. So I did the thing my reasons recommend, but I did it accidentally. This is because what explains why I jumped in (the flying roof tile and/or intention) is completely independent from what explains what makes it rational to jump in (the fact that the child requires aid). Thus, we can explain [my jumping & why it would be rational to jump] by stitching together the explanation for why I jumped and the explanation why it would be rational to jump, which have no bearing on one another. But we can’t explain why I rationally jump. Cases like these falsify what we might call “naïve views of orthonomy” that assume all that needs to hold for an orthonomous relationship is that the agent acts and they have a reason to act. They falsify such (hypothetical) views by exploiting the fact that, if there is no appropriate connection between the components of the presumed orthonomous relationship, then they can be explained independently. Call this philosophical device an *exploit*.

Second, consider the causalist thought that orthonomous relationships are causal connections. We already saw that this suggestion is vitiated by the example of David. David’s action is caused by the fact that provides his reason, but through a process that is deviant relative to the nature of the normative demand. David does what his reasons demand accidentally. This is because under the causalist proposal, only a composite answer to the coincidence question is possible in cases like that of David. The very possibility of deviance is based in the fact that the same proposition might have several explanatory roles that figure in independent explanations. In the case of David, the fact which provides his reason has two independent explanatory roles. As a causal factor, it explains why David breaks up with his partner. As a normative factor, it explains why it would be rational to break up with his partner. But it fails to offer a unifying explanation. It explains why it was rational to break up, and why David broke up with his partner. But it doesn’t explain David’s rationally breaking up with his partner (which is to say that no such action occurs). We have still merely stitched together two independent explanations.

Cases of deviance are explanatory exploits. They falsify causalism by exploiting the fact that the components of the orthonomous relationship can be explained independently by the same factor. This is why causalism about orthonomy must fail—because there is nothing in the repertoire of causal theories par excellence, no purely causal notion that is, that could

guarantee that a given cause doesn't play two independent explanatory roles. Unless this possibility is ruled out, the possibility of cases of deviance will remain.

Third, consider the modalist thought that the orthonomous relationship between reason and action can be understood modally. My counterexample to modalism was the case of Roland, who only does what Daniele (who is very good at responding to reasons) does. Roland does what his reasons demand accidentally. This is because under the modalist proposal, only a composite answer-albeit a very complex composite answer-to the coincidence question is possible in cases like that of Roland. To see this, consider what Roland's modal profile really consists of. Roland has the right kind of modal profile if his reasons and actions are modal companions. Across a relevant modal subspace, they occur together in all (or most) possible situations. Thus, a modal profile is just a quantification over a set of discrete situations (or worlds) in which it is true that $[\varphi \& R]$. It is typically added that reason and action are causally related at the relevant worlds, but as we saw, this fact alone will not provide the relevant relational information. When we ask why $[\varphi \& R]$, then, the modalist response is that all (or most) of the possible R's correspond to all (or most) of the possible φ -ings. But the situations (or worlds) at which these truths hold are, by definition, not related except by relations of similarity. And so for we all know, there might be completely independent and separate explanations for why $[\varphi \& R]$ holds at w_1 , and for why it holds at w_2 , and so on. In this way, it is still possible to explain the components of $[\varphi \& R]$ independently, or in other words, to decompose this fact into a multitude of separate independent facts about R and about φ . In fact, this is why cases of mimicking, masking, and finking are possible in the first place - because nothing guarantees that a set of modal truths that pattern in a certain way don't hold independently of one another. If they do, a modal pattern will not amount to the properties or relations in the actual world it indicates. Cases of masking, mimicking, and finking are therefore modal exploits. They falsify modalism by exploiting the fact that modal patterns may hold without any underlying relational facts in the actual world, and that we can therefore explain the components of orthonomy, modally conceived, independently.

What these escalating ways of exploiting composite answers to coincidence questions show is that *successful* answers to coincidence questions about orthonomy need to be *fully unified*. Recall that since answers to coincidence questions occur within the context of certain explanatory interests, what counts as coincidence relative to one context might not count as a coincidence relative to another. We can express this insight in another way: An explanation of $[p \& q]$ is unifying iff there is *at least one* context according to which we cannot explain p and q independently. An explanation of $[p \& q]$ is *fully* unifying if there is *no* context according to which we can explain p and q independently. The kind of non-accidentality that is required for freedom and orthonomy requires the explanatory relationship to be fully unifying.

Thus, we have come to an informative characterisation of the kind of explanation that dispenses accidentality, and which consequently gives us an explanationist account of the orthonomous relationship that grounds freedom:

(E+) Free agency with respect to S 's φ -ing is grounded in a fully unifying explanation of S 's φ -ing in terms of sufficient normative reasons.

This account of freedom captures our judgments, and their bases, about all of the cases I have considered so far, including our judgements about the Frankfurt Case in section 1. But it also captures what we find important about the concept of free agency more generally.

We care about free agency because it matters to us that we, as agents, are part of the explanation of our actions. This is revealed in cases in which we are not part of the explanation in the right way, when we are the victims of manipulation, to give a prominent example. In cases of manipulation, it is the activity and the reasons of another agent, not our own, which explains why we act. When Silvio manipulates me into breaking up with my partner, then what explains my action are Silvio's reasons, and Silvio's actions, but not my own. And we accordingly judge that actions that are the outcome of manipulation are not free, exactly as (E+) tells us.

More broadly, we care about free agency *inter alia* because we would like to think that our practical lives and their accompanying concepts are not just epiphenomena of a causally closed world. For if this was true, then, to put it dramatically, the fact that our reasons, plans, and intentions tend to match up with the actual events would turn out to be a huge coincidence. What we want is a story about our practical lives in which this threat of coincidence is credibly eliminated (whether this is a compatibilist or an incompatibilist story is another matter). What we want, when we want this story, is to be able to explain the actual events and our actions as one; to synthesise the intentional and the causal descriptions of our doings into one unified picture. At the core of this worry is therefore the very aspect spelled out by (E+), the connection between explanation and accidentality inherent in the concept of free agency.

6. Conclusion

It is plausible that agents are free in virtue of their orthonomy—their relationship to the demands of their reasons. I have argued that neither a causal nor a modal account of this relationship is very promising. I have instead suggested that it is much more natural to understand orthonomy as a kind of explanatory relationship between reason and action. This relationship, I have suggested, obtains when we can give a specific unifying story about how our reasons and actions connect. We are free and basically accountable for our actions, that is, when we can rationally account for them. Or in other words: When we can tell a story about how our actions

are exercises of our rational capacities. This, I take it, was the core of orthonomy accounts all along. Explanationism just offers an improved way of expressing the thought.

References:

- Aguilar, Jesús H. (2012). Basic causal deviance, action repertoires, and reliability. *Philosophical Issues* 22 (1), 1-19.
- Alvarez, M. (2018). Reasons for action, acting for reasons, and rationality. *Synthese* 195 (8), 3293-3310.
- Bishop, J. (1989). *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge University Press.
- Bhogal, Harjit (2020). Coincidences and the Grain of Explanation. *Philosophy and Phenomenological Research* 100 (3), 677-694
- Bogardus, T. and Perrin, W. (2022). Knowledge is Believing Something Because It's True. *Episteme*, 1-19.
- Carroll, L. (1895). What the tortoise said to Achilles. *Mind* , 4(14), 278-280.
- Clarke, R. (2008). Dispositions, Abilities to Act, and Free Will: The New Dispositionalism. *Mind* 118, 323-351.
- Comesaña, J. and McGrath, M. (2014). Having False Reasons. In Clayton Littlejohn and John Turri (eds.), *Epistemic Norms*. Oxford University Press, 59-80.
- Davidson, D. (1963) Actions, Reasons, and Causes. *Journal of Philosophy* 60, 685-700.
- Fara, M. (2008). Masked Abilities and Compatibilism. *Mind*, 117 (468), 843-65.
- Faraci, D. (2019). Groundwork for an Explanationist Account of Epistemic Coincidence. *Philosophers' Imprint*, 19.
- Fischer, J. M. and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, 66, 829-839.
- Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68, 5-20.
- Frankfurt, H. (2003) Some Thoughts Concerning PAP in D. Widerker and M. McKenna (eds) *Moral Responsibility and Alternative Possibilities*, Ashgate.
- Franklin, C. E. (2011). Masks, Abilities and Opportunities: Why the new dispositionalism cannot succeed. *Modern Schoolman*, 88, 89-103.
- Ginet, C. (1996) In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. *Philosophical Perspectives* 10, 403-17.

- Goldman, Alvin I. 1967. "A Causal Theory of Knowing." *Journal of Philosophy* 64 (12): 357-372.
- Heering, D. (forthcoming). Failure and Success in Agency. *Philosophical Quarterly*.
- Heering, D. (2022a). Actual Sequences, Frankfurt-Cases, and Non-accidentality. *Inquiry: An Interdisciplinary Journal of Philosophy* 65 (10):1269-1288.
- Heering, D. (2022b). Reasons-responsiveness, modality and rational blind spots. *Philosophical Studies* 180 (1):293-316.
- Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.
- Hu, J. (2023). Reasons-Responsiveness and the Challenge of Irrelevance. *Journal of the American Philosophical Association* 9 (4):762-778.
- Lando, T. (2017). Coincidence and Common Cause. *Noûs*, 51 (1), 132-151.
- Lange, M. (2010). What Are Mathematical Coincidences? *Mind*, 119 (474), 307-40.
- Lange, M. (2016). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford University Press USA.
- Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge University Press.
- Lord, E. and Sylvan, K. (2019). Prime Time (for the Basing Relation). In Carter, A. J. and Bondy, P. (eds.), *Well Founded Belief: New Essays on the Epistemic Basing Relation*.
- Lutz, M. (2020). Explanationism provides the best explanation of the epistemic significance of peer disagreement. *Philosophical Studies*, 177 (7), 1811-1828.
- Marcus, Eric (2012). *Rational Causation*. Harvard University Press.
- McKenna, M. (2013). Reasons-Responsiveness, Agents, and Mechanisms. In: *Oxford Studies in Agency and Responsibility Volume 1*. Oxford University Press.
- Monod, J. (1970). *Le Hasard et la Necessite: Essai sur la philosophie naturelle de la biologie moderne*. Editions du Seuil.
- Morton, Adam (1975). Because he thought he had insulted him. *Journal of Philosophy* 72 (1):5-15.
- Neta, R. (2019). The Basing Relation. *Philosophical Review* 128 (2):179-217.
- Owens D. (1992). *Causes and Coincidences*. Cambridge University Press.
- Peacocke, C. (1979). Deviant Causal Chains. *Midwest Studies in Philosophy* 4 (1):123-155
- Pettit, P. and Smith, M. (1990). Backgrounding Desire. *Philosophical Review*, 99, 565-592.
- Pettit, P and Smith, M. (1996). Freedom in Belief and Desire. *Journal of Philosophy*, 93, 429-449.
- Sartorio, C. (2011) *Actuality and Responsibility*. *Mind* 120, 1071-97.
- Sartorio, C. (2016). *Causation and Free Will*. Oxford University Press UK.

- Sartorio, C. (2022). The grounds of our freedom. *Inquiry: An Interdisciplinary Journal of Philosophy* 65 (10):1250-1268.
- Schroeder, M. (2008). Having reasons. *Philosophical Studies*, 739(1), 57-71.
- Shope, R. K. (1992). You know what you falsely believe (or: Pollock, know thyself!). *Philosophy and Phenomenological Research*, 52 (2), 405-410.
- Smith, M. (1997). A theory of freedom and responsibility. In Cullity, G. and Gaut, B. (eds.), *Ethics and Practical Reason*. Oxford University Press, 293-317.
- Smith, M. (2003). Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In Stroud, S. and Tappolet, C. (eds.), *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press, 17-38.
- Smith, M. (2004). The Structure of Orthonomy. *Royal Institute of Philosophy Supplement*, 55, 165-193.
- Sosa, E. (2017). *Epistemology*. Princeton: Princeton University Press.
- Sosa, E. (2007). *Apt Belief and Reflective Knowledge, Volume 1: A Virtue Epistemology*, Oxford: Oxford University Press.
- Swenson, P. (2015). A challenge for Frankfurt-style compatibilists. *Philosophical Studies*, 172(5), 1279-1285.
- Van Inwagen, P. (1983) *An Essay on Free Will*, Oxford: Clarendon Press.
- Vihvelin, K. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics*, 32, 427-450.
- Vihvelin, K. (2013). *Causes, Laws, and Free Will*. New York: Oxford University Press.
- Whittle, A. (2010). Dispositional Abilities. *Philosophers' Imprint*, 10(12) (September), 1-23.
- Widerker, D. (2000). Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look. *Philosophical Perspectives*, 14, 181-201.
- Widerker, D. (2003). Blameworthiness, and Frankfurt's Argument Against the Principle of Alternative Possibilities in D. Widerker and M. McKenna (eds) *Moral Responsibility and Alternative Possibilities*, Ashgate.
- Wolf, S. (1987). Sanity and the Metaphysics of Responsibility, in Schoeman 1987: 46-62.
- Wolf, S. (1990). *Freedom Within Reason*. Oup Usa.