Erkenntnis (2025) 90:1541–1559
https://doi.org/10.1007/s10670-023-00760-w

ORIGINAL RESEARCH

Graded Abilities and Action Fragility

David Storrs-Fox[1,2]

Received: 11 August 2022 / Accepted: 19 November 2023 / Published online: 26 December 2023

Abstract
Recent work by Alfred Mele, Romy Jaster and Chandra Sripada recognizes that abilities come in degrees of fallibility. The rough idea is that abilities are often not surefire. They are liable to fail. The more liable an ability is to fail, the more fallible it is. Fallibility is plausibly significant for addiction, responsibility, and normative theorizing. However, we lack an adequate account of what fallibility consists in. This article addresses that problem. Perhaps the most natural approach is to say (roughly) the fallibility of your ability to F is the proportion of scenarios in which you do not F, among those in which you try to F. I argue that this approach (in all plausible versions) is mistaken. I then introduce the notion of an action's "fragility," and propose that we use that new notion to understand fallibility.

1 Introduction

The literature on abilities has usually proceeded as if abilities are all-or-nothing, instead of coming in degrees.[1] The abilities literature shares that feature with literatures on other modal phenomena, like dispositions, possibility, and necessity. However, recent work in philosophy and linguistics has begun to take seriously the thought that all these phenomena are graded.[2] The present article concerns the thought that abilities come in degrees of *fallibility* (Jaster, 2020; Mele, 2003, 2017, 63–89; Sripada, 2018, 2019).[3]

[1] Maier (2014)'s survey article does not mention the thought that abilities come in degrees, and Mandelkern et al. (2017)'s theory of agentive modal terms like 'able' does not consider phrases like 'better able' or 'more (or less) able.'

[2] On dispositions, see Manley and Wasserman (2007, 2008), Vetter (2015). On possibility and necessity, see Lassiter (2017).

[3] 'Fallibility' is Sripada's term. Mele and Jaster discuss the same phenomenon using the term 'reliability.' An ability is fallible just in case it is (in Mele and Jaster's terminology) not perfectly reliable.

✉ David Storrs-Fox
dstorrsfox@gmail.com

[1] Institute for Ethics in AI and Jesus College, University of Oxford, Oxford, UK

[2] Institute of Philosophy, Chinese Academy of Sciences, Beijing, China

We can introduce fallibility using a well-known case from J.L. Austin (1979). Austin supposes he is able to hole a very short putt, yet misses when he tries to hole it. Austin presents his case as a counterexample to the "traditional conditional" account of all-or-nothing abilities, but it also reveals a way in which abilities appear to come in degrees. Abilities are often not sure-fire. They are liable to fail. The more liable an ability is to fail, the more fallible it is.

Suppose Austin is playing with a professional golfer, Bennett. Plausibly, her ability to hole the putt is less fallible than Austin's. That is not to deny that Austin is able to hole the putt. But Bennett is better able to hole it, in that her ability is less liable to fail. An *infallible* ability is sure-fire, not liable to fail. A perfect golfer would have an infallible ability to hole Austin's putt, and (presumably) an omnipotent agent's abilities would all be infallible. (Perhaps actual humans rarely, if ever, have infallible abilities.) At the other extreme, someone is *totally unable* to do something if she has no ability at all to do it, not even a highly fallible one. I might be totally unable to change the past, or prove a contradiction.

Philosophers often distinguish between general and specific abilities (Jaster, 2020; Maier, 2015; Mandelkern et al., 2017; Whittle, 2010). Suppose Rafael Nadal is in bed sick. It is true (in some sense) to say that Nadal is able to serve faster than 110mph, even though (in another sense) his sickness makes him unable to serve at all. As the point is commonly put: Nadal has the *general* ability to serve faster than 110mph, but not the *specific* ability (i.e., the ability on this specific occasion).

Importantly, Austin's claim is that he has the specific ability to hole the putt. It is not that Austin is generally able to hole putts like this, but on the occasion described (like Nadal in bed) he lacked the ability. Rather, he was able on this very occasion to hole the putt. As Austin puts it: "It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it" (Austin, 1979, fn. to 218). And that ability – that *specific* ability – brings with it some liability to failure. I will confine my discussion in this article to specific abilities, remaining neutral on whether and how general abilities are graded.

Austin's judgments are not universally accepted. Some authors deny that Austin had the ability to hole the putt on that specific occasion. They seek to account for the inclination to judge otherwise by saying he possessed merely the *general* ability to hole putts on relevantly similar occasions (e.g. Berofsky, 2011, 155; White, 1968, 87–89). These authors would perhaps deny that specific abilities come in degrees of fallibility. They might say that if you are able to do something on a specific occasion, your ability is sure-fire. However, many others agree with Austin (e.g. Alvarez, 2013, 108–109; Fara, 2008, 847, 862–863; Jaster, 2020, 54–55; Mele, 2003, 449; 2017, 65–66; Thalberg, 2013, 126–129; Vihvelin, 2004, 442).

My aim is to seek an account of what fallibility is: of what it is (for example) for Austin's ability to hole the putt to be more fallible than Bennett's. I will assume (with Austin) that abilities on specific occasions are sometimes fallible, and (with Jaster, Mele and Sripada) that their fallibility comes in degrees. As such, this article does not directly address the dispute about whether Austin has the specific ability to hole the putt. Still: it bears indirectly on that dispute, by providing a theoretical

framework on which Austin *could* miss the putt while having the specific ability to hole it.

But why care about fallibility? Here are two reasons. Firstly, fallibility plausibly bears on the many topics to which abilities are relevant (cf. Mele, 2017, 84–86). Consider addiction. It is a familiar thought in addiction research that addiction impairs some of an addict's abilities without removing them altogether (Holton & Berridge, 2013; Levy, 2010; Sinnott-Armstrong, 2012). The notion of fallibility is plausibly one of the tools required to understand that impairment. For example, Chandra Sripada's Fallibility Model of addiction relies heavily on the thought that our self-control abilities are not infallible (Sripada, 2018).

Consider also the principle that 'ought' implies 'can': that is, that you ought to do something only if you are able to do it. We should not simply ask whether that principle is true. We should also ask how fallible your ability to do something can be, if it is to be a candidate for what you ought to do. Indeed, perhaps the relevance of our abilities to what we ought to do is a matter of degree – so that, for example, you often have less reason to do something the more fallible your ability to do it. Consider finally the Principle of Alternative Possibilities (PAP), according to which you are morally responsible for something you did only if you were able to do otherwise. Again, we should not simply ask whether PAP is true. We should also ask how fallible your ability to do otherwise can be, if you are to be responsible for what you did. And perhaps you are less responsible the more fallible that ability was.

The second reason to care about fallibility is its importance for understanding abilities themselves. A complete theory of abilities should say something about the ways in which they come in degrees. Of course, we might restrict our attention to all-or-nothing abilities – that is, to whatever is ascribed using sentences like 'Austin is able to hole the putt.' But even so, it would be odd if our theory of all-or-nothing abilities were entirely disconnected from the ways in which abilities are graded (cf. Jaster, 2020, 23–28).

The literature on fallibility is in its infancy. To my knowledge, the only rigorous account of the phenomenon is Jaster's (2020). Jaster's account is an example of the *modal tie approach* to fallibility, which accounts for fallibility in terms of a "modal tie" between tryings (or decidings, or intendings) and successful performances.[4] More exactly, the modal tie approach says the fallibility of your ability to *F* is the proportion of the possible situations in which you do not *F*, among those in which you try (or decide, or intend) to *F*.[5] The modal tie approach resembles the "traditional conditional" account of all-or-nothing abilities (TCA), according to which you are able to *F iff* you would *F* were you to try to *F*. TCA falls prey to (what I call) the *Obstruction Problem* (Chisholm, 1964; Davidson, 2001; Lehrer, 1968). However, its more sophisticated descendants – such as the account in Mandelkern et al. (2017) – overcome that problem.

I argue in Sect. 2 that the modal tie approach to fallibility falls prey to the *Graded* Obstruction Problem. Importantly, the maneuver TCA's descendants employ to deal

[4] The phrase 'modal tie' is Jaster's.

[5] Throughout this article, I use '*F*' to stand in for a verb phrase that picks out a type of action or omission.

with the Obstruction Problem does not help with its graded cousin. I conclude that the Graded Obstruction Problem is fatal for the modal tie approach to fallibility. Sections 3, 4 and 5 develop an alternative approach, which understands the fallibility of the ability to *F* in terms of (what I call) the *fragility* of various *F*-ings. Section 3 provides an account of fragility. Section 4 justifies accounting for fallibility indirectly in terms of fragility, and Sect. 5 applies my account of fallibility to examples. The fragility approach provides a promising alternative to the modal tie approach, and does not fall prey to the Graded Obstruction Problem.

My ambitions are in two ways modest. Firstly, I do not seek to provide a complete account of fallibility: I leave many details to be filled out. Secondly, I do not seek to deal exhaustively with possible alternatives to the fragility approach. Having dismissed the modal tie approach, I do not find any other alternatives promising. I leave it to those who disagree to make a case for their preferred alternative. Here I focus on the fragility approach, with the aim of providing a basis for future theorizing about fallibility.

2 The Graded Obstruction Problem

The traditional conditional account (TCA) of all-or-nothing abilities is widely discussed and widely rejected (Berofsky, 2011). The account says you are able to *F iff* you would *F* were you to try to *F*. (Other versions replace trying with deciding or intending. These variations do not matter for present purposes.) Perhaps the most obvious account of fallibility would simply translate TCA to the graded context. For instance, consider the possible situations in which you try to *F*. The translated account might say the fallibility of your ability to *F* is the proportion of those situations in which you do not *F*. The present section argues against that account of fallibility, as well as against related accounts. But first, it will help to review the relevant literature on TCA and all-or-nothing abilities.

It is generally accepted that TCA fails to provide a sufficient condition for the possession of an ability, because someone might be unable even to try (or decide) to do something (Chisholm, 1964, 24–25; Davidson, 2001, 68). Call this the *Obstruction Problem*.[6] Here's an example from Keith Lehrer (1968, 32). Celia has a pathological aversion to red sugar balls, which remind her of blood. Such an aversion might render someone unable even to try (or decide) to take a red sugar ball. Celia would take a red sugar ball were she to try to take one, but because of her aversion she is unable to try. She can take a red sugar ball *only* by trying to take one, so her inability to try makes her unable to take one. TCA falsely entails that Celia is able to take a red sugar ball, so it fails to state a sufficient condition for the possession of an ability.

The Obstruction Problem defeats TCA, but not its more sophisticated descendants. It seems that Celia's example is problematic for TCA because the situation(s) that would be actualized were Celia to take a red sugar ball is one in which she lacks her pathological aversion (or at least: in which her aversion is less powerful

[6] I call it that because in the relevant cases, the agent is obstructed from trying, deciding, etc.

than it actually is). The fact that she would take a red sugar ball in *that* situation seems to have no bearing on whether she is actually able to take one. Indeed, there is a possible situation in which Celia sprouts wings and flies away, but the fact that she would fly in that situation has no bearing on whether she is actually able to fly. An account of abilities should not count situations like this as relevant to an agent's actual abilities. Plausibly, the issue both with the situation in which Celia lacks her (powerful) aversion and the situation in which she sprouts wings is that they differ from the actual situation in respects that matter. As I will put it: these situations are not *realistic*.

Accordingly, Christopher Peacocke's account of abilities supplements TCA's counterfactual conditional with the requirement that there be a close possible situation in which the agent tries to *F*. On Peacocke's view, possible situations in which Celia lacks the aversion will not count as close. Because there is no close situation in which Celia tries to take a red sugar ball, Peacocke's account correctly says she is unable to take one (Peacocke, 1999, 311–314).[7]

Mandelkern et al. (2017) provide another variant of TCA. On their view, you are able to *F iff* there is some "practically available action" such that if you were to try to do it, you would *F* (314). Taking a red sugar ball is not a practically available action for Celia, because of her aversion. Consequently, possible situations in which she tries to take a red sugar ball are not counted as relevant to whether she is able to take one. There is plausibly no other practically available action such that if Celia were to try to do *it*, she would take a red sugar ball. So Mandelkern et al.'s account correctly says Celia is unable to take one (Mandelkern et al., 2017, 317–323).

On Peacocke's view, which situations are close has to do with whether we can rely on their not obtaining (Peacocke, 1999, 310–312). On Mandelkern et al.'s view, which actions are practically available typically has to do with what the agent could reasonably conclude in favor of doing (Mandelkern et al., 2017, 318–325). Neither of these approaches can be applied mechanically to ascertain which possible situations are relevant for an agent's abilities. However, each provides the materials for excluding from relevance situations that are not (in my terminology) realistic – like those in which Celia lacks her aversion.

With this background in mind, here is a natural way of translating TCA into an account of fallibility.

> Modal Tie: Count as relevant the realistic possible situations in which you try to *F*. The fallibility of your ability to *F* is the proportion of the relevant situations in which you do not *F*. If there are no relevant situations, you are totally unable to *F*.

The talk of proportions in Modal Tie is to be applied as follows. If you *F* in all the relevant situations, you are infallibly able to *F*. If you *F* in none of the relevant situations, you are totally unable to *F*. And the greater the proportion of relevant situations in which you do not *F*, the more fallible your ability to *F*. Modal Tie will likely require making sense of proportions among infinitely many possible

[7] Peacocke talks about "freedom" rather than "abilities," and of "possibilities" rather than "possible situations." These differences are unimportant for my purposes.

situations. I agree with others that this can be done, although the task is far from straightforward (Bigelow, 1976; Jaster, 2020, 173–175; Manley & Wasserman, 2008, 82). For the present article, I set it aside.

As with TCA, Modal Tie has variants that involve deciding or intending in place of trying. Jaster's account is one such variant: she takes the relevant possible situations to be the realistic situations in which you intend to *F* (Jaster, 2020, 95, 103–104).[8] My comments concerning Modal Tie will apply also to these variants.

I am not aware of any detailed account of fallibility in the literature other than Jaster's. Mele (2017, 63–89) discusses fallibility using the term 'reliability', but gives no account of the phenomenon. Still, Modal Tie is a natural extension of the dispositional analyses of abilities provided by Fara (2008), Vihvelin (2013) and Sosa (2015) (cf. Jaster, 2020, 188–199).

Let us consider two versions of the Obstruction Problem in relation to Modal Tie. The account deals easily with the first. Suppose Daniel shares Celia's aversion. He is guaranteed to take a red sugar ball if he tries to take one: that is, he faces no further liability to failure *given* that he tries. However, because of his aversion he is totally unable even to try. He can take a red sugar ball only by trying to take one, so his inability to try renders him totally unable to take one. Someone might object to Modal Tie that all the possible situations it counts as relevant are situations in which Daniel tries to take a red sugar ball – but in all those situations, he takes one. Therefore (someone might think), Modal Tie falsely entails that Daniel is infallibly able to take a red sugar ball.

However, Modal Tie counts only *realistic* situations as relevant. Its proponent can therefore apply the same maneuver as Peacocke and Mandelkern et al. The account simply needs to exclude from the realistic situations any situation in which Daniel tries to take a red sugar ball. It could borrow from Peacocke, and say there is no close possibility in which Daniel tries to take a red sugar ball. Or it could borrow from Mandelkern et al., and deny that taking a sugar ball is a practically available action for Daniel. Jaster employs similar machinery in her account of fallibility (Jaster, 2020, 97–98, 108–135). Whichever approach we take, there are no realistic possible situations in which Daniel takes a red sugar ball. Modal Tie then correctly says Daniel is totally unable to take a red sugar ball.

So far so good for Modal Tie. But the account faces a second (and more troubling) version of Lehrer's problem. Call it the *Graded Obstruction Problem*. In summary, the problem is that an ability to *try* might itself be fallible. Modal Tie is sensitive only to situations in which the agent tries to *F*, because it looks at the proportion of *those* situations in which the agent *F*-s. But this is too narrow: there are possible situations in which the agent does not manage even to try to *F*, and which affect the fallibility of her ability to *F*. The appeal to realisticness allows Modal Tie to handle a *total* inability to try (as in Daniel's case), but not a somewhat fallible ability to try.

[8] Jaster later allows that sometimes, the intention to perform some action other than *F*-ing can play the role she assigns to the intention to *F* (2020, 140–153). This complication is not relevant to my concerns.

The account I have described is Jaster's account of abilities to perform intentional actions ("agentive abilities"). She also extends her view to cover "non-agentive" abilities, like the abilities to digest and see (2020, 178). I will focus on abilities to perform intentional actions, so will not discuss the extended view.

To develop the Graded Obstruction Problem, consider that nearly everything we do (including mental actions) is implemented in part by an extremely complex sequence of cognitive steps. Many of these steps are carried out by subpersonal systems over which we have little control, and which are subject to error. Errors might arise because of the stochasticity of neural processes, the misallocation of limited cognitive resources, or other factors (Sripada, 2018, 12–14; cf. Redish et al., 2008). Such subpersonal errors can cause agents to fail in higher-level tasks. For example, a subpersonal error could cause a tennis player to hit the ball slightly too hard and miss a serve (Sripada, 2018, 9–14).

Return to the action of taking a red sugar ball. An error that causes the agent *not* to take a ball could occur anywhere in the cognitive sequence that implements the action. Such an error could occur relatively late in the sequence, so that the steps up to that point are sufficient for the agent to have decided, intended, and tried to take the red sugar ball. Alternatively, it could occur earlier in the sequence. Mental actions like deciding and (if different) forming an intention themselves involve a complex cognitive sequence (Sripada, 2018, 10–11). Subpersonal error could interrupt such a sequence before events sufficient for (say) a decision have occurred, by (for example) causing the agent to become confused or distracted (cf. Sripada, 2018, 13–14).

The same is presumably true of trying. There is considerable disagreement about what it is to try to do something (like take a red sugar ball). (Ginet, 1990; Hornsby, 1980; Jones, 1983; O'Shaughnessy, 1973, 2007, 598–608; Ruben, 2018; Sharvit, 2003; Yaffe, 2010, esp. 92–94.) But whatever trying is, it will surely involve a complex sequence of cognitive steps that are themselves subject to error. An error that occurs sufficiently early in the sequence could prevent the agent even from trying (e.g. through distraction).[9]

Suppose Emma has no aversion to red sugar balls. Like Celia and Daniel, she can take a red sugar ball only by trying to take it. As with nearly all our actions, the sequence of cognitive steps that could implement her taking a red sugar ball is susceptible to error. Let us suppose – unrealistically, to make the point clearer – that the (small[10]) susceptibility to error is present only in the cognitive steps required for her to try to take a red sugar ball. Because of this susceptibility to error, she is not infallibly able to try to take a red sugar ball. But as long as she does try, she is guaranteed to take one.

Emma's case then poses a problem for Modal Tie. Emma is able with a low degree of fallibility to take a red sugar ball, given that the susceptibility to error is small. But she is not *infallibly* able to take one, because there is some risk that she will not manage even to try to take one (and so will not take one). However, Modal Tie cannot yield this in-between judgment about Emma's abilities: it will either say

[9] Indeed, many philosophers hold that trying to do something *requires* deciding or forming the intention to do it (Ginet, 1990; Hornsby, 1980; Jones, 1983; O'Shaughnessy, 2007, 598–608; Yaffe, 2010, 92–94).

[10] The susceptibility to error needn't be small. Sripada (2018) suggests that fatigue and psychosocial stressors could make subpersonal cognitive errors more likely (16). More fancifully, nefarious surgical intervention could produce the same result (cf. Frankfurt, 1969).

that she is totally unable able to take a ball, or else that she is infallibly able to take one.

This time, the maneuver from Peacocke, Mandelkern et al. and Jaster will not work. To avoid the claim that Emma is totally unable to take a red sugar ball, Modal Tie needs to count as realistic *some* of the possible situations in which she tries to take one. But whichever situations in which she tries are counted as realistic, all of them will be situations in which she takes a red sugar ball. Modal Tie will then falsely entail that Emma is infallibly able to take a red sugar ball. The same points apply to the variants of Modal Tie (like Jaster's) that put *deciding* or *intending* in place of trying, and even to variants that put *trying* to try or *beginning* to try in place of trying. We are not infallibly able to do any of these things, because they are all implemented by complex and error-prone cognitive sequences. I conclude that the Graded Obstruction Problem will arise for any plausible variant of Modal Tie.

Of course, we do not need the unrealistic supposition that Emma is guaranteed to take a red sugar ball if she tries. If (more realistically) there is some susceptibility to error throughout the cognitive sequence that implements Emma's action, Modal Tie will no longer say Emma is infallibly able to take a red sugar ball: there will be some situations in which she tries to take a red sugar ball but does not. Still, the account will *understate* the fallibility of Emma's ability, because it will be insensitive to errors that prevent her even from trying to take a ball.

Jaster acknowledges that the ability to intend might itself be fallible. However, she does not see that observation as a problem for her account of abilities to perform intentional actions. After all (she says), intending is not itself an intentional action (Jaster, 2020, 227). The challenge she sees is simply that of giving an account of the ability to intend (2020, 227–228). But the Graded Obstruction Problem presents a more serious challenge for Jaster's view, and for other variants of Modal Tie. Because of Emma's fallible ability to intend or try to take a red sugar ball, she is also more fallibly able to *take* a red sugar ball. The Graded Obstruction Problem challenges Modal Tie as an account of the fallibility of abilities to perform intentional actions.

If we reject Modal Tie on the basis of the Graded Obstruction Problem, we might wonder whether a more complicated account could accommodate its core feature. A natural suggestion is that the fallibility of the ability to F depends on the proportion Modal Tie invokes *together with* the fallibility of your ability to try (or intend, etc.) to F.

This suggestion shifts the problem to the account of the ability to try (or intend, etc.). It will not do to apply Modal Tie to *that* ability, so that we count as relevant the realistic possible situations in which you *try* to try to F. The Graded Obstruction Problem will apply here too, because (as I argued above) we are not infallibly able to try to try to F. We might instead opt for a hybrid account, combining the proportion in Modal Tie with a different approach to the ability to try or intend. Jaster herself (tentatively) suggests a view on which the fallibility of your ability to intend to F is the proportion of situations in which you do not intend to F, among those in which you have "overriding reason" to intend to F (Jaster, 2020, 227–228). (Here Jaster is discussing only the ability to intend to F; she does not advocate a hybrid account of the ability to F.)

However, this view faces problems. There will surely be realistic possible situations in which agents do not do things that they have overriding reason to do. Suppose I have overriding reason to intend to eat my spinach. I might still not intend to eat it, simply because I do not like spinach. My non-intending need not reflect any degree of inability so to intend. In light of such a situation, Jaster's proposal will count me as more fallibly able to intend to eat my spinach. But that is the wrong result: plausibly, an agent could even be *infallibly* able to form some intention, and yet irrationally not form it (cf. Davidson, 2001, 77; Kaiserman, 2021, 695–696).

Jaster might reply that if I have *truly* overriding reason to intend to eat the spinach and yet do not intend to eat it, it must be because of (partial) inability that I do not so intend. If your reason to intend to eat the spinach is *truly* overriding (she might reply), it is impossible for you freely to defy it.

But even if there are such overriding reasons, we surely lack them in many cases. Suppose then that I have good (but not overriding) reason to intend to eat my spinach, and my ability to intend to eat it is nearly infallible. Jaster's proposal directs us to the counterfactual situations in which I *do* have overriding reason to intend to eat my spinach. It might be that in most of *those* situations, I would "choke" under pressure and not manage to form the intention (Beilock, 2007). Jaster's proposal will then incorrectly say that my ability to intend to eat my spinach is highly fallible. The problem is that in a case in which someone lacks overriding reason to (intend to) do something, it might be that her ability to intend would be different if she *were* to have overriding reason.

These problems will afflict the hybrid account of the ability to *F*, if it incorporates Jaster's account of the ability to intend to *F*. The hybrid account will overstate the fallibility of my ability to eat the greens in the first variant of the case, and understate it in the second variant. The hybrid theorist might seek a different account of the ability to try or intend to *F*. However, finding such an account is no trivial task. It will also be reasonable to ask whether, given an account of the ability to try to *F*, we could simply apply that same account to the ability to *F* itself. It is not obvious that the hybrid structure will be needed once we have an account of the ability to try.

In my view, there is no plausible variant of Modal Tie that can overcome the Graded Obstruction Problem. I am also pessimistic about the prospects for a hybrid account. The Graded Obstruction Problem provides good reason to seek a different approach to fallibility: one that gives no special role to tryings (or intendings, etc.). The rest of this article develops such an approach.

3 Fragility

My approach to fallibility involves what I call *fragility*. Fragility is a feature not directly of the ability to *F*, but of an *F*-ing: that is, of a performance (i.e., a token) of the action type picked out by '*F*'. The rough idea is that when an agent does something she is only very fallibly able to do, she is in some sense lucky to have done the thing. I intend the notion of fragility to capture the relevant sort of luckiness.

Roughly: the more fragile an *F*-ing, the luckier (in the relevant sense) the agent is to have *F*-ed.[11]

Over the next three sections, I will develop and defend the following account of fallibility:

> Fragility Min: Take the fragility of your *F*-ing in each of the realistic possible situations in which you *F*. The fallibility of your ability to *F* is the lowest of these fragility values. If there are no realistic possible situations in which you *F*, you are totally unable to *F*.

The present section proposes an account of fragility. Section 4 justifies accounting for fallibility indirectly (via fragility), as well as Fragility Min's appeal to the *lowest* fragility value. Section 5 illustrates Fragility Min using examples.

Consider again the example of Austin playing golf with Bennett, a professional golfer. They face the same short putt, and are both able to hole it. However, Bennett's ability to hole the putt is significantly less fallible than Austin's. Let us suppose that Austin's ability is highly fallible, whereas Bennett's is nearly infallible. Now suppose that both Austin and Bennett hole the putt. Still, Austin was in some way very lucky to hole the putt: much luckier than Bennett was.

The relevant sort of luck is bound up with fallibility. Agents who are very fallibly able to *F* are (in the relevant way) lucky when they manage to *F*. Our actions are subject to other sorts of luck that are *not* closely connected with fallibility. Suppose Bennett was nearly hit by a car on her way to the golf course. Perhaps she would not even have arrived at the course had she been hit, or perhaps she would have arrived with two broken arms. In the actual situation (where she was not hit), she is perhaps lucky to hole the putt rather than (say) being in hospital. But we can also make sense of the thought that her holing of the putt was *not* especially lucky, and that Austin's was luckier, given the respective fallibility of their abilities. It is this fallibility-related luck I intend fragility to capture. Austin's holing of the putt was very fragile, whereas Bennett's holing was not especially fragile.

Consider next the example of Emma, which I used to present the Graded Obstruction Problem. Emma is not infallibly able to take the red sugar ball, because there is some risk that subpersonal cognitive error will prevent her from taking it. Importantly, in Emma's case such error would prevent her even from deciding or trying to take a red sugar ball. To alter the case slightly, let us suppose that Emma is *very* susceptible to relevant cognitive error. Consequently, she is very fallibly able to take a red sugar ball. She is also very fallibly able even to try (or decide, etc.) to take a red sugar ball. In the event, Emma manages to take a ball. Just as Austin was lucky to hole the putt, Emma was lucky to take a ball. There was a significant risk that subpersonal cognitive error would prevent her from taking a ball, and also prevent her even from trying (or deciding, etc.) to take one. It seems that Emma's taking of

[11] Kikkert (2022) relates her notion of "local control" to luck in performance in a similar way: the luckier (in some sense) a performance is, the lower the degree to which the agent has local control over that performance. I will remain neutral here on how local control relates to fragility.

the ball is fragile, like Austin's holing of the putt. The way in which she was lucky to have taken the ball is bound up with fallibility.

I think of fragility as analogous to safety – or rather, unsafety – in epistemology (Pritchard, 2005; Sosa, 1999; cf. Kikkert, 2022). Actions can be fragile in something like the way true beliefs can be unsafe. Just as an unsafe true belief is in some way lucky, a fragile action is in some way lucky. And in both cases, the luck is plausibly connected with whether relevant risks are manifested in close possible situations: that is, in situations that could easily have been actualized. I am inclined to think there are extensive and instructive analogies between the subject matter of epistemology and that of action theory (Anscombe, 1963; Williamson, 2017, 2018). However, I mention the analogy with safety here merely for expository purposes, not to justify my account of fragility.

I suggest that we understand the fragility of an *F*-ing in terms of the proportion of close possible situations that contain relevant non-*F*-ings. Which non-*F*-ings are relevant? Consider Austin's holing of the putt. Suppose that in some close situation, Austin tries his best to hole the putt but does not manage. Plausibly, this non-holing is relevant for fragility. That is, a theory of fragility should count Austin's holing of the putt as more fragile in light of it. However, some relevant non-*F*-ings will not involve the agent even trying to *F*. Consider Emma's taking of the red sugar ball. Suppose that in some close possible situation, subpersonal cognitive error prevents Emma even from trying to take a red sugar ball, and so from taking one. Plausibly, this possible non-taking of a ball is relevant for fragility: a theory of fragility should count Emma's taking of the ball as more fragile in light of it.

There are other possible non-*F*-ings that a theory of fragility should *not* count as relevant. Consider a possible situation in which Bennett intentionally misses the putt, in order to make Austin feel better. Perhaps this situation could easily have occurred: although Bennett actually holed the putt, she pitied Austin *nearly* enough to miss intentionally. Even so, Bennett's non-holing in this situation does not reflect any lack of ability to hole the putt, and should not lead a theory of fragility to count her actual holing as more fragile.

I suggest that the relevant non-*F*-ings are the *unintentional* non-*F*-ings. (I call an *F*-ing 'unintentional' just in case it is not intentional.) In a situation where Austin tries his best but still misses, his non-holing is unintentional. In a situation where subpersonal cognitive error prevents Emma even from trying to take a sugar ball (e.g., through distraction), her non-taking of the ball is unintentional. By contrast, in a situation where Bennett misses to make Austin feel better, her non-holing is intentional. Austin's non-holing and Emma's non-taking could therefore bear on a performance's fragility according to my account, whereas Bennett's non-holing could not.

Here is the crucial point that will, in a moment, allow Fragility Min to deal with the Graded Obstruction Problem: an agent can unintentionally not *F* without even trying, deciding, or intending to *F*. To make the point vivid: suppose an explosion destroys my brain five seconds before I would have sipped my coffee, and three seconds before I would even have thought about sipping. I do not sip the coffee, and my non-sipping is not intentional. I also do not try, decide, or intend to take a sip.

I take it that a performance could be intentional or unintentional only *as* a performance of some action type. For example, your hitting the bullseye might be intentional as a hitting of the dartboard, but not as a hitting of the bullseye. You intentionally hit the dartboard, but you did not intentionally hit the bullseye.[12] [13] Similarly, a performance could be fragile or unfragile only *as* a performance of some action type. Your hitting the bullseye might be fragile *as* a hitting of the bullseye, but not as a hitting of the dartboard. Throughout this article, when I say an *F*-ing is (un)intentional or fragile I am saying it is (un)intentional or fragile *as* an *F*-ing.

Here, then, is my proposal. Consider a situation in which you *F*. Take all the possible situations close to that one, then remove from them any situations in which you intentionally do not *F*. (The latter sort of situation is irrelevant for fragility.) The fragility of your *F*-ing is the proportion of the remaining situations in which you do not *F*. That is: the numerator situations (as it were) are the close situations in which you unintentionally do not *F*. The denominator situations are all the close situations *apart from* those in which you intentionally do not *F* (i.e., the close situations in which either you *F* or you unintentionally do not *F*). As I noted above for Modal Tie, I assume we can make sense of proportions among infinitely many possible situations (cf. Bigelow, 1976).

I will not say much here about which situations are close for the purposes of fragility (i.e., which situations could easily have been realized). As with safety in epistemology, I take it this will be a complex and context-sensitive matter (Williamson, 2009). However, it is worth recalling the point that actions might be subject to luck of a sort that is not relevant to fragility or fallibility. Bennett was in some way lucky to hole the putt because she was nearly hit by a car on her way over, but her holing was not any more fragile for that reason. In my view, that is because the causal process leading to her non-holing in this counterfactual situation (in which she is hit by a car) is sufficiently different from the causal process leading to her actual holing (cf. Dunaway & Hawthorne, 2017, 291–292).[14]

By contrast, consider a possible situation in which subpersonal processes implementing her action cause her to hit the ball slightly too hard, and she misses the putt. The causal process leading to *this* non-holing is significantly more similar to the process leading to her actual holing, and is plausibly sufficiently similar to be relevant to the holing's fragility.

We can incorporate this point into an account of fragility via the relation of closeness between possible situations (cf. Williamson, 2009, 20–21). That is: we count a

[12] Philosophers sometimes make the same point by saying actions are (un)intentional "under a description" (Anscombe, 1963, 37–39).

[13] Here I assume a coarse-grained view of action individuation, on which your (token) bullseye-hitting is identical with your (token) dartboard-hitting. If you doubt that identity claim, you can substitute my example for another in which a performance falls under multiple action types. If you think there are no such examples, you can think of performances as (un)intentional and (un)fragile *simpliciter* (not simply *as* performances of some action type).

[14] It is plausible that in the counterfactual situation in which Bennett is hit by the car, her non-holing of the putt is unintentional. But because this counterfactual situation is not close to the actual situation (in which she holes the putt), it is still not relevant to the fragility of Bennett's actual holing.

situation as close only if it involves an *F*-ing or non-*F*-ing with a sufficiently similar etiology as the *F*-ing being evaluated for fragility.

4 All-Or-Nothing Ability Ascriptions

The present section justifies Fragility Min's indirect approach: that of understanding a graded feature of abilities (i.e., fallibility) in terms of a graded feature of performances (i.e., fragility). It also argues that the *least* fragile performance is what matters for fallibility.

To justify Fragility Min's indirect approach, let us turn to all-or-nothing ability ascriptions: that is, to utterances of sentences like 'Austin is able to hole the putt.' Such ascriptions belong to a larger class of utterances involving modal terms like 'can' and 'able.' Many of these utterances do not concern abilities (e.g., 'The food here can be overcooked,' 'You cannot drive faster than 15mph near a school'). An account of fallibility should fit with our best theory such utterances (cf. Jaster, 2020, 23–28). Semantic orthodoxy understands modal terms like 'can' and 'able' in terms of existential quantification: an utterance of the form 'A can *F*' or 'A is able to *F*' is true *iff* A *F*-s in some relevant possible situation. There is ample linguistic evidence for this orthodoxy, across multiple languages (Kratzer, 2012, 1–69).

Let a *direct view* of fallibility be one on which the fallibility of your ability to *F* is the proportion of situations in which you do not *F*, among the relevant possible situations. A direct view appeals to such a proportion directly in its account of fallibility, rather than indirectly via something like fragility. Modal Tie is a direct view, as is Jaster's variant of it. For the former, the relevant situations are the realistic situations in which you try to *F*. For the latter, they are the realistic situations in which you intend to *F*.

Direct views of fallibility fit uneasily with an orthodox semantics of all-or-nothing ability ascriptions. Plausibly, an ability-ascribing utterance of the form 'A can *F*' or 'A is able to *F*' is true only if the ability is not too fallible. Suppose I want to shoot a pigeon on a distant roof. I am not *totally* unable to hit the pigeon, but my ability to hit it is highly fallible. It would (in many contexts) be false to say, 'I can hit the pigeon [from here]' or 'I am able to hit the pigeon [from here].'

It seems that such utterances are true only if the ability does not exceed some threshold of fallibility. That threshold will plausibly be vague, and depend on the context of utterance (Jaster, 2020, 26–28). Applying a direct view of fallibility, we reach the view that the utterances are true only if the proportion of relevant situations in which I do not hit the pigeon is not above some threshold. That will usually require that, among the many relevant possible situations, there be substantially more than one situation in which I *do* hit the pigeon. If there is only one such situation, my ability to hit the pigeon will usually be too fallible to make the utterance true. However, on the orthodox semantics for 'can' and 'able' – which simply involves existential quantification – a single relevant situation in which I hit the pigeon is enough for the utterance to be true. Direct views are therefore in tension with the orthodox semantics.

To solve this problem, we can draw on an observation made independently by Wolfgang Schwarz (2020) and Malte Willer (2021): that the context of utterance might restrict *what counts* as an *F*-ing, for an ability ascription of the form 'A can *F*' or 'A is able to *F*.' As Schwarz and Willer show, this observation elegantly solves various ability-related problems for the orthodox semantics. Applying the observation to the pigeon example, we can say that the utterance of 'I can hit the pigeon' is false because there is no possible situation in which I do something that counts (in the context of utterance) as a pigeon-hitting. It seems we should therefore understand the fallibility of my ability to hit the pigeon in terms of some feature of the various pigeon-hittings (i.e., of *performances*).

I nominate fragility as that feature. On my view, a pigeon-hitting must not exceed some threshold of fragility if it is to count for an ability ascription. (Again, the threshold will plausibly be vague and context-dependent.) It is false to say 'I can hit the pigeon' or 'I am able to hit the pigeon,' because there is no sufficiently unfragile pigeon-hitting. I might luckily hit the pigeon, but then my pigeon-hitting would be too fragile to count. The all-or-nothing ability ascriptions are true if, and only if, there is at least one sufficiently fragile pigeon-hitting.

This is an advantage of understanding fallibility indirectly in terms of fragility. But why does Fragility Min identify an ability's fallibility with the *lowest* of the relevant fragility values? On the suggested elaboration of the orthodox semantics, the utterance of 'I can hit the pigeon' is true only if there is *some* relevant pigeon-hitting whose fragility does not exceed the threshold. That is what it takes for the ability not to exceed its fallibility threshold. There is some such relevant pigeon-hitting if, and only if, the fragility of the *least fragile* relevant pigeon-hitting does not exceed the threshold. If the fragility of the least fragile hitting *does* exceed the threshold, so does the fragility of every other hitting. And if it does not, that is enough for there to be *some* hitting whose fragility does not exceed the threshold.[15]

For the ability not to exceed its fallibility threshold, it is necessary and sufficient that the least fragile performance not exceed its fragility threshold. In my view, the simplest way to accommodate this point in a theory of fallibility is to *identify* the ability's fallibility with the fragility of the least fragile pigeon-hitting. An existential quantification account of all-or-nothing ability ascriptions supports a minimal fragility view of fallibility.

5 A Fragility-Based Account of Fallibility

Here is my account of fallibility again:

[15] A complication involves cases with infinitely many relevant *F*-ings, such that for every *F*-ing there is another that is less fragile. Such cases (if possible) have no least fragile *F*-ing. To extend Fragility to these cases, we can say the fallibility of your ability to *F* is such that for each relevant *F*-ing, the fallibility of your ability to *F* is no greater than the fragility of that *F*-ing. The account would then provide upper bounds for the fallibility of your ability, but would not (in these cases) deliver an exact fallibility value. In my view, we should simply say in such cases that it is vague exactly how fallible your ability is.

> Fragility Min: Take the fragility of your *F*-ing in each of the realistic possible situations in which you *F*. The fallibility of your ability to *F* is the lowest of these fragility values. If there are no realistic possible situations in which you *F*, you are totally unable to *F*.

(The relevant fragility values might be tied for lowest.)

Before applying Fragility Min to examples, let me comment on its appeal to *realistic* possible situations. I mentioned realisticness in Sect. 2, when discussing TCA. The point is that an account of abilities should not count as relevant possible situations that differ from actuality in respects that matter. There are possible situations where Celia lacks her aversion to red sugar balls, and where she sprouts wings and flies away – but these situations are not relevant to what she is actually able to do. They are therefore counted as (in the relevant sense) unrealistic.

Importantly, we usually should *not* hold facts about the agent's desires or intentions fixed across realistic situations (Berofsky, 2005, 196–197; Jaster, 2020, 119–120). I have no desire to smash my computer screen right now, and it would be unlike me to do so. Still, we should not hold fixed this fact about my desires when evaluating my abilities. Many possible situations in which I desire to smash the screen (and do so) are relevant to whether I am able to smash the screen, and so should count as realistic. (This case illustrates how the present use of 'realistic' deviates from ordinary English. In *some* sense, the situations in which I smash my screen are highly unrealistic – I wouldn't do such a thing – but that sense does not concern us here.)

Which situations count as realistic plausibly depends on the context of ability-ascribing utterances (cf. Jaster, 2020, 108–135; Mandelkern et al., 2017, 318–325). A detailed theory of realisticness would require considerably more than one article, and I will not provide one here. Instead, I simply emphasize that the need for such a theory will likely arise for any plausible account of abilities, including Modal Tie and TCA's descendants.

I turn now to some examples. Let us begin with Emma, and the Graded Obstruction Problem her case illustrates. Recall that Emma can take a red sugar ball only by trying to take one. Because of her susceptibility to subpersonal cognitive error, she is not infallibly able even to try to take a red sugar ball. Consequently, she is not infallibly able to take one. Modal Tie struggles with Emma's case because it is sensitive only to possible situations in which she *tries* to take a red sugar ball, ignoring those in which she does not manage even to try.

Fragility Min can do better. Let a *try-error* be a subpersonal error that prevents Emma even from trying to take a red sugar ball (e.g., through distraction). Fragility Min directs us to the realistic situation with the least fragile ball-taking. Among the situations close to that one, there will plausibly be situations containing try-errors. (There may not be *many* such try-error situations, but her susceptibility to try-errors should lead us to expect *some*.) In the situations containing try-errors, Emma does not take a sugar ball, and her non-taking is unintentional. On my account of fragility, these situations therefore increase the fragility of the least fragile ball-taking. So according to Fragility Min, they increase the fallibility of Emma's ability to take a red sugar ball (the correct result). That is so even though she does not *try* to take a

ball in the try-error situations. Again, the key point is that an agent can unintentionally not *F* without trying to *F* (or deciding or intending to *F*).

Here is another case that helps to illustrate Fragility Min. Suppose that Austin has two ways he can hole the putt: via a trick shot, or via an ordinary shot. He is able with a low degree of fallibility to hole it via the ordinary shot, but very fallibly able to hole it via the trick shot. Applying Fragility Min, these fallibility claims are true just in case the least fragile ordinary-shot holing (in a realistic situation) is not especially fragile, while the least fragile trick-shot holing is very fragile.

Importantly, Fragility Min will also say in this case that the fallibility of Austin's ability to hole the putt is simply the fallibility of his ability to hole it via an ordinary shot. That is because the *least* fragile putt-holing is a putt-holing via an ordinary shot. This seems plausible, even leaving aside the semantics of all-or-nothing abilities. Suppose Austin initially has only the ability to hole the putt via the ordinary shot, which is not especially fallible. He then gains the highly fallible ability to hole the putt via the trick shot. His ability to hole the putt surely does not for that reason become *more* fallible. Similarly, my ability to drink my coffee without spilling is not especially fallible. That is so even though my ability to do that while holding the mug two feet above my mouth *is* very fallible.

Section 4 justified Fragility Min's appeal to the least fragile performance, and I take examples like these to support that appeal further. However, it is worth making two further points for those who wonder whether other (more fragile) performances are sometimes relevant for fallibility.

Firstly, abilities are plausibly graded in ways that are orthogonal to fallibility. When Austin learns the trick shot he becomes (we might say) more *flexibly* able to hole the putt, in that he is now able to hole it in more ways than he could before (cf. Jaster, 2020, 128–135; Kittle, 2015). An ability's flexibility plausibly is sensitive to a range of performances in different realistic possible situations. However, I maintain that Fragility Min delivers the correct verdict concerning the *fallibility* of Austin's ability to hole the putt. Gaining the highly fallible ability to hole the putt via the trick shot does not make Austin any more (or less) fallibly able to hole the putt, given that he already has a not-especially-fallible ability to hole it in the ordinary way.

Secondly, we should keep in mind the Schwarz-Willer observation that the context in which an ability to *F* is ascribed might affect what counts as an *F*-ing, for the purposes of that ascription. A hitting of the bullseye from two inches away might be less fragile than any bullseye-hitting that conforms to the rules of darts. But in ordinary contexts of utterance the former (illegal) bullseye-hitting will not count as a bullseye-hitting, when discussing someone's ability to hit the bullseye. It will therefore not count as the least fragile bullseye-hitting.

The Schwarz-Willer observation also helps with an example from an anonymous reviewer. Suppose (because of visual impairment) my ability to read a letter in a normal way is very fallible. However, my ability to read it while holding it close to my eyes is not especially fallible. Still, we might judge that I am very fallibly able to read the letter. The reviewer proposes that the fallibility of the ability to *F* is a matter of the fragility of *F*-ings done in a normal way. I doubt this proposal when applied to other cases. Suppose Austin is very fallibly able to hole the putt, then learns a highly

abnormal (but perfectly legal) technique. His ability to hole the putt via the new technique is nearly infallible. In my view, he is now nearly infallibly able to hole the putt. But the reviewer's proposal would (incorrectly) treat holings via the new technique as irrelevant to the fallibility of his ability to hole the putt.

Still, Fragility Min can use the Schwarz-Willer observation to accommodate the reviewer's judgment concerning the letter example. The idea would be that in relevant contexts of utterance, letter-readings conducted in abnormal ways do not count as letter-readings for the ability ascription. The fallibility of my ability to read the letter will then be the fragility of the least fragile letter-reading that is conducted in a normal way.

6 Conclusion

There remain details to fill in. Fragility Min needs a theory of exactly which possible situations are *realistic*, relative to the situation in which the ability is possessed. It also needs a theory of exactly which possible situations are *close*, relative to a situation in which the ability is exercised. In my view, the details of both will depend on the context of the relevant ability ascriptions. It might also be that to make judgments about realisticness and closeness, we will need to draw on our judgments about abilities themselves (cf. Williamson, 2009, 9–10).

I have also not discussed examples of "impairment by ignorance."[16] Suppose I have no idea what the code to your safe is. I am nearly infallibly able to enter '83463', which happens to be the code. It seems that there could be both true and false readings of the sentence 'I am nearly infallibly able to open the safe.' On the false readings, it seems my ignorance impairs my ability. In common with other accounts of abilities and fallibility, a complete fragility-based account should capture these false readings (Carlson, 1999; Jaster, 2020, 146–149; Schwarz, 2020;).

I offer Fragility Min not as the final word on fallibility, but as a promising basis for future work. Given the potential significance of fallibility – for the theory of abilities, for addiction, for what we ought to do, and for responsibility – that work is worth doing.

7 Competing Interests

The author has no relevant financial or non-financial interests to disclose.

[16] This term is Jaster's.

References

Alvarez, M. (2013). Agency and two-way powers. *Proceedings of the Aristotelian Society, 113*(1), 101–121.
Anscombe, G. E. M. (1963). *Intention*. 2nd ed.
Austin, J. L. (1979). Ifs and Cans. In *Philosophical Papers* (3rd ed., pp. 205–232). Oxford University Press.
Beilock, S. L. (2007). Choking under pressure. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 140–141). SAGE Publications.
Berofsky, B. (2005). Ifs, cans, and free will: The issues. In R. Kane (Ed.), *The Oxford handbook of free will* (1st ed., pp. 181–201). Oxford University Press.
Berofsky, B. (2011). Compatibilism without Frankfurt: Dispositional analyses of free will. In R. Kane (Ed.), *The Oxford handbook of free will* (2nd ed., pp. 153–174). Oxford University Press.
Bigelow, J. C. (1976). Possible worlds foundations for probability. *Journal of Philosophical Logic, 5*(3), 299–320.
Carlson, E. (1999). The oughts and cans of objective consequentialism. *Utilitas, 11*(1), 91–96.
Chisholm, R. M. (1964). J.L. Austin's philosophical papers. *Mind, 73*(289), 1–26.
Davidson, D. (2001). Freedom to act. In *Essays on actions and events* (pp. 63–82). Oxford University Press.
Dunaway, W., & Hawthorne, J. (2017). Scepticism. In W. J. Abraham & F. D. Aquino (Eds.), *The Oxford handbook of the epistemology of theology* (pp. 290–308). Oxford University Press.
Fara, M. (2008). Masked abilities and compatibilism. *Mind, 117*, 843–865.
Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy, 66*(23), 829–839.
Ginet, C. (1990). *On action*. Cambridge University Press.
Holton, R., & Berridge, K. (2013). Addiction between compulsion and choice. In N. Levy (Ed.), *Addiction and self-control* (pp. 239–268). Oxford University Press.
Hornsby, J. (1980). *Actions*. Routledge & Kegan Paul Ltd.
Jaster, R. (2020). *Agents' abilities*. De Gruyter.
Jones, O. R. (1983). Trying. *Mind, 92*, 368–385.
Kaiserman, A. (2021). Reasons-sensitivity and degrees of free will. *Philosophy and Phenomenological Research, 103*.
Kikkert, S. (2022). Ability's two dimensions of robustness. *Proceedings of the Aristotelian Society, 122*(3), 348–357.
Kittle, S. (2015). Abilities to do otherwise. *Philosophical Studies, 172*, 3017–3035.
Kratzer, A. (2012). *Modals and conditionals*. Oxford University Press.
Lassiter, D. (2017). *Graded modality*. Oxford University Press.
Lehrer, K. (1968). Cans without Ifs. *Analysis, 29*(1), 29–32.
Levy, N. (2010). Addiction and compulsion. In T. O'Connor & C. Sandis (Eds.), *A companion to the philosophy of action* (pp. 267–273). Wiley-Blackwell.
Maier, J. (2014). Abilities. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy,* Fall 2014 Edition. URL: http://plato.stanford.edu/archives/fall2014/entries/abilities/
Maier, J. (2015). The agentive modalities. *Philosophy and Phenomenological Research, 90*(1), 113–134.
Mandelkern, M., Schultheis, G., & Boylan, D. (2017). Agentive modals. *Philosophical Review, 126*(3), 301–343.
Manley, D., & Wasserman, R. (2007). A Gradable approach to dispositions. *The Philosophical Quarterly, 57*(226), 68–75.
Manley, D., & Wasserman, R. (2008). On linking dispositions and conditionals. *Mind, 117*(465), 59–84.
Mele, A. R. (2003). Agents' abilities. *Noûs, 37*, 447–470.
Mele, A. R. (2017). *Aspects of agency*. Oxford University Press.

O'Shaughnessy, B. (1973). Trying (as the mental "Pineal Gland"). *Journal of Philosophy, 70*(13), 365–386.
O'Shaughnessy, B. (2007). *The will: A dual aspect theory* (Vol. 2). Cambridge University Press.
Peacocke, C. (1999). *Being known*. Oxford University Press.
Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.
Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral and Brain Sciences, 31*(4):415–437; discussion 437–487.
Ruben, D.-H. (2018). *The metaphysics of action: Trying, doing, causing*. Palgrave Macmillan.
Schwarz, W. (2020). Ability and possibility. *Philosophers' Imprint, 20*(6), 1–21.
Sharvit, Y. (2003). Trying to be progressive: The extensionality of Try. *Journal of Semantics, 20*(4), 403–445.
Sinnott-Armstrong, W. (2012). Are addicts responsible? In N. Levy (Ed.), *Addiction and self-control* (pp. 122–143). Oxford University Press.
Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives, 13*, 141–153.
Sosa, E. (2015). *Judgment and agency*. Oxford University Press.
Sripada, C. (2018). Addiction and fallibility. *The Journal of Philosophy, 115*, 1–19.
Sripada, C. (2019). The fallibility paradox. *Social Philosophy & Policy, 36*(1), 234–248.
Thalberg, I. (2013). *Enigmas of agency*. Routledge.
Vetter, B. (2015). *Potentiality: From dispositions to modality*. Oxford University Press.
Vihvelin, K. (2004). Free will demystified: A dispositional account. *Philosophical Topics, 32*(1&2), 427–450.
Vihvelin, K. (2013). *Causes, laws, and free will: Why determinism doesn't matter*. Oxford University Press.
White, M. (1968). On what could have happened. *The Philosophical Review, 77*(1), 73–89.
Whittle, A. (2010). "Dispositional Abilities." *Philosophers' Imprint, 10*(12), 1–23.
Willer, M. (2021). Two puzzles about ability can. *Linguistics and Philosophy, 44*, 551–586.
Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy, 4*, 1–35.
Williamson, T. (2017). Acting on knowledge. In J. A. Carter, E. C. Gordon, & B. W. Jarvis (Eds.), *Knowledge first: Approaches in epistemology and mind* (pp. 163–181). Oxford University Press.
Williamson, T. (2018). Knowledge, action, and the factive turn. In V. Mitova (Ed.), *The factive turn in epistemology* (pp. 125–141). Cambridge University Press.
Yaffe, G. (2010). *Attempts*. Oxford University Press.